This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C12N 15/31, C07K 14/255, A61K 39/112, C12N 1/20 // (C12N 1/20, C12R 1:42)

A2

(11) International Publication Number: WO 94/25598

(43) International Publication Date: 10 November 1994 (10.11.94)

(21) International Application Number: PCT/IB94/00207

(22) International Filing Date: 26 April 1994 (26.04.94)

(30) Priority Data:
08/054,452 26 April 1993 (26.04.93) US

(71) Applicants: UNIVERSITY OF VICTORIA INNOVATION AND DEVELOPMENT CORPORATION [CA/CA]; University of Victoria, P.O. Box 1700, Victoria, Bristish Columbia V8W 2Y2 (CA). KING, Joshua [US/US]; 7002 - 127th Place SE, Renton, WA 98056 (US).

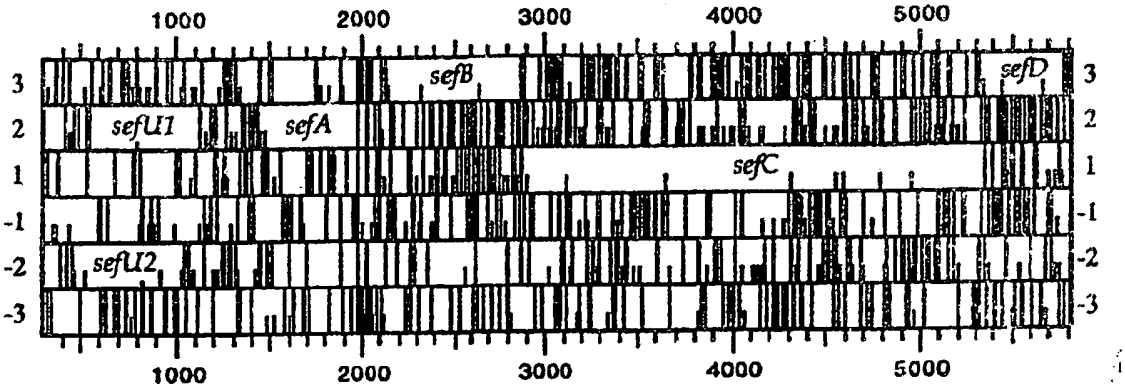
(72) Inventors: KAY, William, W.; 1608 Hampshire Road, Victoria, British Columbia V8R 5T5 (CA). COLLINSON, S., Karen; 7025 Hagan Road, R.R. #1, Brentwood Bay, British Columbia V0S 1A0 (CA). CLOUTHIER, Sharon, C.; 5965 Broadway Road, Nanaimo, British Columbia V9V 1C9 (CA). DORAN, James, L.; 7025 Hagan Road, R.R. #1, Brentwood Bay, British Columbia V0S 1A0 (CA).

(74) Common Representatives: KING, Joshua et al.; Seed and Berry, 6300 Columbia Center, 701 Fifth Avenue, Seattle, WA 98104-7092 (US).

(81) Designated States: AU, BB, BG, BR, BY, CA, CN, CZ, FI, GE, HU, JP, KG, KP, KR, KZ, LK, LV, MD, MG, MN, MW, NO, NZ, PL, RO, RU, SD, SI, SK, TJ, TT, UA, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: METHODS AND COMPOSITIONS FOR *SALMONELLA*-BASED VACCINES



(57) Abstract

Isolated nucleic acid molecules comprising one or more of the *sefB*, *sefC*, *sefD*, *sefU*$_1$, *sefU*$_2$, *agfA*, *tctA*, *tctB*, or *tctC* genes of *Salmonella*. Isolated proteins encoded by said genes. Methods and compositions for eliciting an immune response in animals utilizing the isolated genes and/or proteins, including the utilization of attenuated *Salmonella*, *E. coli*, *Shigella* and other bacteria produced pursuant to induced mutuations in certain of the described genes.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**GB**	United Kingdom	**MR**	Mauritania
AU	Australia	**GE**	Georgia	**MW**	Malawi
BB	Barbados	**GN**	Guinea	**NE**	Niger
BE	Belgium	**GR**	Greece	**NL**	Netherlands
BF	Burkina Faso	**HU**	Hungary	**NO**	Norway
BG	Bulgaria	**IE**	Ireland	**NZ**	New Zealand
BJ	Benin	**IT**	Italy	**PL**	Poland
BR	Brazil	**JP**	Japan	**PT**	Portugal
BY	Belarus	**KE**	Kenya	**RO**	Romania
CA	Canada	**KG**	Kyrgystan	**RU**	Russian Federation
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SI**	Slovenia
CI	Côte d'Ivoire	**KZ**	Kazakhstan	**SK**	Slovakia
CM	Cameroon	**LI**	Liechtenstein	**SN**	Senegal
CN	China	**LK**	Sri Lanka	**TD**	Chad
CS	Czechoslovakia	**LU**	Luxembourg	**TG**	Togo
CZ	Czech Republic	**LV**	Latvia	**TJ**	Tajikistan
DE	Germany	**MC**	Monaco	**TT**	Trinidad and Tobago
DK	Denmark	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**US**	United States of America
FI	Finland	**ML**	Mali	**UZ**	Uzbekistan
FR	France	**MN**	Mongolia	**VN**	Viet Nam
GA	Gabon				

Description

METHODS AND COMPOSITIONS FOR *SALMONELLA*-BASED VACCINES

Cross-Reference To Related Applications

This application is a continuation-in-part of United States patent application Serial No. 08/054,452, filed April 26, 1993 and presently pending.

Technical Field

The present invention relates generally to methods and compositions for eliciting an immune response, and more particularly to such compositions directed to, or based on, *Salmonella*.

Background of the Invention

In 1980, the World Health Organization estimated that in developing countries food poisoning from infection with *Salmonella* bacteria (salmonellosis) contributed in a major way to food borne infections resulting in more than 1 billion cases of acute diarrhea in children under the age of five years (Kvenberg and Archer, *Food Technol. 40*:77-98, 1987, and at least 5 million deaths (this reference, and all other references cited herein, is hereby expressly incorporated herein by reference in its entirety). Since the mid-1980s, the worldwide incidence of salmonellosis has increased steadily. *S. enteritidis* (also known as *S. enterica* ser. *Enteritidis*), in particular, has been implicated in the sharp increase in food borne infection in industrialized countries since 1983. Indeed, the current frequency of *S. enteritidis* infections is considered to constitute a worldwide pandemic (Rodrigue et al., *Epidemiol. Infect. 105*:21-27, 1990).

The severity of the disease is greatest in infants, the elderly, the infirm and in other persons with inadequate or impaired immune systems, including the malnourished. In third world countries where malnutrition is more commonly a complicating factor, mortality rates due to *S. enteritidis* infection as high as 28% have been reported. In both the clinical and industrial settings, the situation is also complicated by the fact that many people are asymptomatic carriers. *Salmonella* spp., including *S. enteritidis*, often possess several plasmid encoded antibiotic resistance genes that complicate the treatment of human infections.

In the industrialized world, it is the contamination of food products by *Salmonella* bacteria that is most directly threatening to human health. Hence, it is not surprising that the increase in salmonellosis in first world countries parallels the

centralization of food production and processing despite continued improvements in epidemiological and microbiological methods.

The significance of the problem is reflected in one aspect in the poultry-related industries. For example, in the U.S. alone hatcheries produce approximately 100 million broiler chicks per week and chicken egg production in the U.S. has reached 5 billion annually. A large proportion of *S. enteritidis* infections have been associated with the contamination of the contents of whole shell eggs resulting from vertical transmission of this pathogen due to transovarian infection. This is significant since common procedures designed to decontaminate the external shell surface are not effective. The problem presented by *S. enteritidis* is exacerbated by the fact that infection in the adult laying hens may be asymptomatic. Typically, *S. enteritidis* infection of laying birds does not have a significant adverse effect on fertility, hatchability or egg production. Similarly, broiler chickens may be asymptomatic throughout their lifetime, although losses of about 20% do occur in infected flocks due to death in chicks, retardation of growth and rejection of contaminated birds at slaughtering. Contaminated poultry feed may be a major source of infection, but stress to poultry due to handling, transportation and overcrowding add to the problem by increasing the shedding of *Salmonella* from infected chickens. The end result is that the majority of modern processing plants, which process about 10,000 birds per hour, are contaminated and *Salmonella* are typically isolated from 40% to 70% of turkey or chicken carcasses sampled in the U.S. and Canada (Lammerding et al. *J. Food Protection 51*:47, 1986).

The overall economic costs of the rising incidence of food borne infections have been significant. The U.S. General Accounting Office has recently estimated the cost of *S. enteritidis* food poisoning in the U.S. between 1985 to 1990 at $118 million in lost productivity, medical and hospital costs resulting from about 9,500 illnesses. The U.S. Center for Disease Control receives more than 40,000 case reports annually but attributes greater than 2 million cases and roughly 2,000 deaths per year in the United States to salmonellosis (Cohen and Tauxe, *Science 234*:964-969, 1986). The economic cost related to treatment of salmonellosis in the U.S. was estimated to be $50 million in 1986. About 8 million cases involve physician consultation and an estimated 250,000 cases require hospitalization. Non-hospitalized cases are thought to have accounted for about $680 million in medical costs and minimally $2 billion in lost productivity. Others estimate the total costs of salmonellosis in the U.S. arising from medical treatment and lost productivity to be as high as $23 billion per year (Kvenberg and Archer, *supra*).

The losses absorbed by the food industry from liability and product loss are undoubtedly passed on to the consumer. Thus, there is a need for an effective risk-management program to monitor the different phases of poultry production including breeding, raising, slaughtering, packing and further processing, distribution and preparation, and consumption. The development of strategies for creating *Salmonella*-free feed, the control of *Salmonella* in breeder flocks, hatcheries, and product operations will include development of more effective diagnostics and vaccines.

It is generally accepted that killed *Salmonella* vaccines are inferior to live attenuated *Salmonella* vaccines so this method has been more-or-less abandoned as a vaccine strategy. Several reports exist that explore humoral response of the host to *Salmonella* components but this strategy has not been widely adopted. Some examples include an outer membrane protein preparation with 4% LPS content, which reportedly gave some protection to *S. typhimurium* in mice (Isibasi et al., *Infect. Immun. 56*:2953, 1988). However, *Salmonella* fimbriae have not been explored for use as subunit vaccines.

Putative vaccines have been developed with *S. typhi* to prevent typhoid fever in humans. However, several non-typhoid *Salmonella* spp. which cause gastroenteritis in humans are host-adapted to cause a typhoid-like disease in certain animals. Consequently, attenuating mutations that have been made in one strain have to be transferred to other *Salmonella* strains to create attenuated mutant vaccines responsive to host-specific diseases. In terms of vaccine design, four basic approaches have been attempted, namely killed vaccine strains, subunit vaccines using purified cell components to elicit protective antibody response, live attenuated strains and live attenuated strains expressing foreign proteins.

Several putative vaccines have been devised that express foreign epitopes on the surface of cells of attenuated *Salmonella*. Some examples include immunogenic epitopes of hepatitis B surface antigens (Wu et al., *Proc. Natl. Acad. Sci. USA 86*:4726-4730, 1989; Schödel et al., in *Progress in Hepatitis B Immunization*, Coursget and Tong, eds., pp. 43-50, 1990), the major surface protein, gp63, of *Leishmania* (Young et al., *J. Immunol. 145*:2281-2285, 1990), the *Streptococcal* M protein (Poirer et al., *J. Exp. Med. 168*:25-32, 1988), and the 31 KDa protein of *Brucella abortus* (Statel et al., *Infect. Immun. 58*:2048-2055, 1990). In some instances, foreign epitopes have been presented on flagella through incorporation of foreign DNA into the flagellin protein gene encoded on an expression vector. For example, a cholera toxin epitope has been expressed from a recombinant plasmid in a *aroA*$^-$, flagellin$^-$ strain of *S. dublin* (Newton et al., *Science 244*:70-72, 1989). However, none of these systems include the use of fimbriae or fimbrial gene products in the vaccines.

A live attenuated *Salmonella* vaccine that persists long enough in the host to elicit a long lasting high level immunity is desired. Several attenuating mutations have been identified that have been used in one or more *Salmonella* spp. singly or in combination for potential use as vaccines. Such vaccines have occasionally shown some protection in mice.

However, there has gone unmet a need for vaccine technologies to reduce the problems associated with food borne, and other, salmonellosis. The present invention provides these and other related advantages.

Summary of the Invention

In one aspect, the present invention provides a composition capable of eliciting an immune response, and preferably a vaccine, comprising an isolated AgfA protein in combination with a physiologically acceptable carrier or diluent. In alternative embodiments, the composition comprises a SefB, SefC, SefD, TctC, TctB or TctA protein.

In a further aspect, the present invention provides vector constructs comprising one or more of a mutant *tctA* gene, *tctB* gene, *tctC* gene or the *tctI* operon that is able to inactivate the corresponding tricarboxylic acid transport pathway in *Salmonella*, to yield an attenuated *Salmonella*. Preferably, the vector construct further comprises a mutant gene that inactivates the ability of *Salmonella* to utilize succinate. Alternatively, the vector construct comprising the mutant *tctA*, *tctB*, or *tctC* gene is used in combination with a vector construct that inactivates the ability of *Salmonella* to utilize succinate. This aspect of the invention therefore provides a biologically pure, attenuated *Salmonella* comprising a mutation in its *tctA*, *tctB*, or *tctC* gene, and/or a gene required for succinate utilization.

This aspect of the invention further provides a composition capable of eliciting an immune response, and preferably a vaccine, comprising an attenuated *Salmonella* that has an inactivating mutation in one or more of its *tctA*, *tctB*, or *tctC* genes in combination with a physiologically acceptable carrier or diluent. Alternatively, the composition comprises an attenuated *Salmonella* having an inactivating mutation in one or more of its *tctI* operon (*tctDCBA*), *tctII* operon, or *tctIII* operon. As a further alternative, the *Salmonella*, either in combination with an attenuating mutation as above or instead of such a mutation, has a mutation in two or more fimbrin encoding genes, such as *sefA*, *sefD*, *afgA or fimA*, wherein the mutation effectively prevents production of fimbriae from such genes.

In preferred embodiments, the attenuated *Salmonella* is able to express a foreign antigen in one or more of its fimbriae. Alternatively, the foreign antigen is fused

to a SefA, SefD, SefC, TctA or AgfA protein. The present invention also provides such attenuated *Salmonella* in a biologically pure form.

In yet a further aspect, the present invention provides an expression vector construct comprising an *agfA* gene that is operably fused in an open reading frame to a foreign gene to yield a dicistronic gene product, the dicistronic gene product able to be expressed in a fimbria of a *Salmonella* or an aggregate comprising such gene product. By "foreign gene" it is meant a gene that is not typically expressed in open reading frame immediately after the *agfA* gene, and preferably a gene that is not found in *Salmonella* in the wild state. Such a foreign gene preferably provides a foreign antigen or epitope, including an immunologically active fragment thereof. In alternative embodiments, the expression vector construct comprises a *tctA* gene, *tctB* gene, *tctC* gene, *sefA* gene, *sefB* gene, *sefC* gene or an *sefD* gene. In a further alternative embodiment, the fimbria or aggregate is expressed in an *E. coli*, or in a *Shigella* spp. In still another alternative embodiment, the expression vector construct comprises an *agfA* gene, *sefA* gene or a *sefD* gene able to produce in *E. coli*, *Shigella* or other *Enterobacteriaceae* a stable fimbria that comprises AgfA protein and/or SefD protein.

In still a further aspect, the present invention provides a stable fimbria or aggregate comprising an AgfA, SefA, SefD, SefC, SefB, TctC, TctB or TctA protein fused to one or more foreign antigens. In an alternative embodiment, this aspect of the present invention provides a stable amino acid polymer comprising an AgfA, SefA, SefD, SefC, SefB, TctC, TctB or TctA protein fused to one or more foreign antigens. Within the context of the present invention, such a stable amino acid polymer, preferably formed in part of AgfA, is able to pass as a naked protein through the stomach of an animal without significant degradation.

In yet a further aspect, the present invention provides methods of eliciting an immune response in, and preferably vaccinating, an animal. In this aspect, fimbriae or aggregates comprising an AgfA, SefA, SefD, SefC, SefB, TctC, TctB or TctA protein fused to a foreign antigen are separated from their *Salmonella* host cell, and then introduced into the animal in conjunction with a physiologically, or pharmaceutically, acceptable carrier or diluent. In an alternative embodiment, the host cell is used to grow a stable amino acid polymer comprising one or more of an AgfA, SefA, SefD, SefC, SefB, TctC, TctB or TctA protein fused to a foreign antigen, which polymer may not comprise a fimbria. In a further alternative embodiment, the fimbriae or aggregate are separated from the host cell prior to being introduced into the animal.

In still a further alternative embodiment, the fimbria or amino acid polymer comprises an AgfA, SefA, SefD, SefC, SefB, TctC, TctB or TctA protein and are grown in an *E. coli* host cell or a *Shigella* host cell. Alternatively, the host cell for

the expression vector construct is *Citrobacter*, *Enterobacteria*, *Pseudomonas*, *Streptomyces*, *Bacillus*, or *Staphylococcus aureus*. Preferably, the AgfA, SefA, SefD, SefC, SefB, TctC, TctB or TctA protein of these alternative embodiments is further fused to a foreign antigen.

Preferably, the fimbriae or amino acid polymers are introduced orally or via injection to the animal. Further preferably, the fimbriae or amino acid polymers are packaged in protein or biodegradable polymers or copolymers, such as polylactide, polygalactide, polycaprolactone, polyanhydride, polyorthoesters, to form microspheres, or are bound to a substrate protein or bacterium. The fimbriae or amino acid polymers may also be maintained in an array.

In yet a further aspect, the present invention provides a method of eliciting an immune response in, and preferably vaccinating, an animal comprising introducing an isolated AgfA protein in combination with a physiologically acceptable carrier into the animal. Preferably, the animal for this method of vaccination, and all other methods of vaccination as described herein, is a human being. In alternative preferred embodiments, the animal is a warm-blooded animal, and further preferably a commercially important warm-blooded animal, including a fowl, a pig, a horse, a dog, a cat, or a cow. In further alternative embodiments, the animal is another commercially important animal such as a shellfish. In alternative embodiments, the isolated protein is a SefD, SefC, SefB, TctC, TctB or a TctA protein.

In still yet another aspect, the present invention provides methods of eliciting an immune response in, and preferably vaccinating, an animal comprising introducing attenuated *Salmonella* into the animal. The attenuated *Salmonella* may comprise one or more of an ineffective *tctA* gene, *tctB* gene, *tctC* gene, *tctI* operon, *tctII* operon, or *tctIII* operon, in one embodiment in combination with mutations of the *agfA*, *sefB*, *sefD* or *fimA*. As noted above, the attenuated *Salmonella* may alternatively comprise two or more ineffective fimbriae genes. In preferred embodiments, the attenuated *Salmonella* expresses one or more foreign antigens. In a still further preferred embodiment, the foreign antigen is located on a SEF17 or SEF18 fimbria of the attenuated *Salmonella*.

It is a further feature of this aspect of the invention that the immune response of the animal can be induced by introducing an *E. coli* into the animal, wherein the *E. coli* expresses a *Salmonella* fimbria. Preferably, the *Salmonella* fimbria comprises an *agfA* gene product, a *sefA* gene product or a *sefD* gene product. Further preferably, the *Salmonella* fimbria further comprises a foreign antigen. Alternatively, a *Shigella* is used in place of the *E. coli*. Further alternatively, another *Enterobacteriaceae* is used in place of *E. coli*.

In yet a further aspect, the present invention provides a method of eliciting an immune response in, and preferably vaccinating, an animal wherein a nucleic acid vector is introduced into the animal. The vector comprises a *sefA* gene, *sefD* gene, *sefC* gene, *sefB* gene, *tctC* gene, *tctB* gene, *tctA* gene, or *agfA* gene. Further preferably, the vector is naked and is injected into a muscle of the animal and/or the above-listed gene is fused to a further, antigen-producing gene to provide an antigenic fusion gene product.

These and other aspects of the present invention will become evident upon reference to the following detailed description and attached drawings. In addition, various references are set forth below which describe in more detail certain procedures or compositions (*e.g.*, plasmids, etc.); such references are incorporated by reference in their entirety.

Brief Description of the Drawings

Figure 1 is a schematic illustration and open reading frame analysis of the *sefU$_2$U$_1$ABCD* gene cluster.

Figures 2A-2D depict the nucleotide sequence of *sefA*, *sefB*, *sefC* and *sefD*, and the corresponding predicted amino acid sequences.

Figures 3A-3B depict the nucleotide sequence of *sefU$_1$* and *sefU$_2$*, and the corresponding predicted amino acid sequences.

Figures 4A-4B depict the nucleotide sequence of *tctA*, and the corresponding predicted amino acid sequence.

Figure 5 depicts the nucleotide sequence of *tctB*, and the corresponding predicted amino acid sequence.

Figures 6A-6B depict the nucleotide sequence of *tctC*, and the corresponding predicted amino acid sequence.

Figure 7A depicts the nucleotide sequence of an *agfA* gene fragment amplified from *S. enteritis* 27655-3b Tn*phoA* mutant strain and cloned into pUC19, and the corresponding predicted amino acid sequence. The solid arrows indicate PCR primer pairs TAF3 and TAF4; the dashed arrows indicate TAF5 and TAF6.

Figure 7B depicts the nucleotide sequence of *agfA*, and the corresponding predicted amino acid sequence.

Figure 8A depicts an autoradiograph of the results of expression of the *sefA*, *sefB* and *sefC* genes in an *E. coli in vitro* transcripti n-translation system. Lane 1, pTZ19; Lane 2, pKX1; Lane 3, pSC1; Lane 4, delB15; Lane 5, delB23; Lane 6, delD10; Lane 7, Western blot of the *in vitro* transcription-translation of pKX1 developed using

antisera generated against denatured SEF14 fimbrin. The size of the molecular weight markers is indicated on the left (10^3 M_r).

Figure 8B is a schematic representation of the *sef* gene cluster showing the inserts of various deletion subclones used in the *in vitro* transcription-translation experiments.

Figure 9 depicts an autoradiograph of the mapping of the 5' end of the *sefA* transcript using primer extension. The lane labeled "*S. enteritidis*" represents the reverse transcriptase products of RNA isolated from this organism grown in CFA static broth for 60 hours at 37°C. The lanes C, A, T, and G represent the results of dideoxy chain termination sequence reactions in the region encompassing the promoter. The sequence of the γ-^{32}P-labeled primer (TGCGTGGGCACTGCCACA) (SEQ ID No. _____) is complementary to nucleotides 181-198 of *sefA*. The arrows indicate two major transcription initiation sites.

Figures 10A-C depict immunoelectron microscopy of negatively stained cells for SEF14 production. (A) *S. enteritidis* 27655-3b; (B) *E. coli* HB101 carrying cos48; (C) *E. coli* JM109 carrying pKX1. Magnification: (A) x115,000; (B) x94,000; (C) x144,000.

Detailed Description of the Invention

The present invention provides methods and compositions for eliciting an immune response from vaccines to *Salmonella*. These methods and compositions include numerous isolated genes specific to *Salmonella*, vector constructs, numerous isolated proteins specific to *Salmonella* and vaccines. These methods and compositions are described further, below.

I. Genes Specific to *Salmonella*

A. Genes Generally

The present invention provides isolated DNA molecules comprising the *sefU$_2$U$_1$ABCD* gene cluster, the *sefABCD* gene cluster, *sefBCD* gene cluster, *sefU$_2$U$_1$* gene cluster, the *sefA* gene, the *sefB* gene, the *sefD* gene, the *agfA* gene, the *tctCBA* gene cluster, the *tctA* gene, the *tctB* gene, and/or the *tctC* gene. Although one embodiment of each of these molecules is shown in Figures 2 to 7B, it should be understood that within the context of the present invention, reference to one or more of these genes includes derivatives of the genes that are substantially similar to the genes (and, where appropriate, the protein (including peptides and polypeptides) that are encoded by the genes and their derivatives). As used herein, a nucleotide sequence is

deemed to be "substantially similar" if: (a) the DNA sequence is derived from the coding region of a native gene of any *Salmonella* serovar and maintains substantially the same biological activity (including, for example, portions of the sequence or allelic variations of the sequences discussed above); (b) the DNA sequence is capable of hybridization to DNA sequences of the present invention under moderate, high or very high stringency (*see* Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2d ed., Cold Spring Harbor Laboratory Press, NY, 1989); or (c) the DNA sequences are degenerate as a result of the genetic code to the DNA sequences defined in (a) or (b). Within the context of the present invention, moderate stringency means conditions such that an appropriate nucleotide sequence is able to selectively hybridize to nucleotide sequences from *Salmonella* and to conserved nucleotide sequences in other bacteria such as *E. coli*. High stringency means the nucleotide sequence is able to selectively hybridize to nucleotide sequences from *Salmonella*. Very high stringency means the nucleotide sequence is able to selectively hybridize to a single *Salmonella* species, such as *S. enteritidis*, *S. dublin* or *S. gallinarum*.

B. *sefA*, *sefB*, *sefC*, *sefD*, *sefU*$_1$, *sefU*$_2$ genes and the *sefU*$_2$*U*$_1$*ABCD* operon

As can be seen by reference to Figure 1, which is a schematic illustration and open reading frame analysis of the *sefU*$_2$*U*$_1$*ABCD* gene cluster, the *sefA*, *sefB*, and *sefC* genes comprise an approximately 3.9 kilobase pair region of *Salmonella* DNA. *In vitro* expression directed by the 5.3 kilobase pair DNA fragment of Figure 1 indicated that the SefA, SefB and SefC proteins have an approximately 14 K, 28 K, and 90 K M_r molecular weight, respectively. *See* Figure 8A. The present invention involves one or more of a *sefA*, *sefB*, *sefC*, *sefD*, *sefU*$_1$, or *sefU*$_2$ gene or gene cluster. Further information with respect to these genes and their products may be found in U.S. Application Ser. No. 08/054,542. *See also* Application Ser. No. ____ (attorney's docket no. 920043.403C1), and Application Ser. No. ____ (attorney's docket no. 920043.403C2). As noted above, these applications and all other references cited herein are expressly incorporated by reference herein in their entirety.

Experimental results with *S. enteritidis* indicate that *sefB* and *sefC* are not expressed in the absence of *sefA*. Primer extension analysis of *sefABC* gene clusters revealed two major transcription start cites located upstream of *sefA* (Figure 9). Transcription of *sefB* and *sefC* is also initiated from the *sefA* promoter region. Secondary structure analysis of the mRNA transcript from *sefABC* predicted the formation of two stable stem-loop structures in the intercistronic region between *sefA* and *sefB*, which is indicative of differential regulation of *sefA* as opposed to *sefB* and

sefC translation. The nucleotide sequences, and corresponding amino acid sequences, of *sefA*, *sefB*, *sefC*, *sefD*, $sefU_1$ and $sefU_2$ are depicted in Figures 2 and 3.

$sefU_1$ and $sefU_2$ are overlapping open reading frames oriented in the opposite directions (*i.e.*, encoded on opposite strands). *sefD* abuts *sefC* and a consensus Shine-Dalgarno ribosomal binding site sequence is just inside the $sefC_{orf}$. No promoter is recognized immediately upstream of *sefD*; expression may be directed by the *sefA* promoter.

C. *agfA* gene

The *agfA* gene codes for a structural fimbrin protein composing very thin fimbriae, approximately 3-4 nanometers in diameter, that are highly aggregative and stable. The aggregative property of the fimbriae is believed to be due at least in part to its hydrophobicity. The gene product of the *agfA* gene, AgfA, is found in SEF17 fimbriae, and comprises an approximately 14-15 K M_r molecular weight fimbrin protein. AgfA contributes to heavy pellicle formation in static cultures, colony hydrophobicity, and autoaggregation of cells in culture. SEF17 fimbriae are immunologically distinct from SEF14 and SEF21, as indicated by the lack of cross-reactivity with polyclonal antisera raised in rabbits against SEF14 or SEF21 (Collinson et al., "Purification and characterization of thin, aggregative fimbriae from *S. enteritidis*," *J. Bacteriol.* *173*:4773-4781, 1991). Figure 7A depicts the nucleotide sequence of an *agfA* gene fragment that was amplified from the *S. enteritidis* 27655-3b Tn*phoA* mutant strain and then cloned into pUC19. The bases underlined in the *agfA* sequence of Figure 7A are common to portions of the PCR primers TAF1 and TAF2, which were used in amplification of this fragment. Below the *agfA* sequence in Figure 7A is the translated amino acid sequence. As discussed further below, the nucleotide sequences targeted by the diagnostic PCR primer pairs TAF3 and TAF4 (solid arrows) and TAF5 and TAF6 (dashed arrows) are also indicated in Figures 3A-3B. Below the *agfA* sequence in Figure 7 is the translated amino acid sequence.

Figure 7B depicts the nucleotide and amino acid sequences of the full *agfA* gene of *S. enteritidis* 27655-3b.

D. *tctA*, *tctB*, and *tctC* genes, and the *tctI*, *tctII* and *tctIII* operons

The *tctA*, *tctB*, *tctC* and *tctD* genes are located in the *tctI* operon, which is one of three operons for tricarboxylic acid transport within *Salmonella* spp. The other *Salmonella* tricarboxylic acid transport systems are termed *tctII* and *tctIII*. It is believed that the tricarboxylic acid transport system of *Salmonella* is lacking from several other related genera of *Enterobacteriaceae*. *tctI* and *tctIII* both encode proteins

responsible for citrate transport and are induced by growth of *Salmonella typhimurium* on minimal media supplemented with citrate as a carbon source. *tctII* is normally not expressed. Each system demonstrates preferential transport of various tricarboxylic acids and different concentrations of the monovalent cations Na^+ and K^+. The transport of tricarboxylic acids, particularly citrate, are potentially important to the intracellular survival of *Salmonella*, as such transport provides a means of scavenging nutrients from a host cell. Accordingly, mutants in these various *tct* systems, preferably in *S. typhimurium* or *S. enteritidis*, potentially create attenuated strains of *Salmonella* capable of being taken into a host cell, and capable of persisting for a limited time, yet not able to proliferate within such a host cell. In a preferred embodiment, one or more of these *tct* mutants are coupled with a mutant unable to utilize succinate to provide a "back-up" system to assure attenuation.

The nucleotide sequences for *tctA*, *tctB* and *tctC*, along with their corresponding amino acids, are depicted in Figures 4, 5 and 6, respectively.

II. Vector Constructs Comprising the Gene Sequences of the Present Invention

A. Vector Constructs Generally

The present invention provides for the manipulation and expression of the above described genes by culturing host cells containing a construct capable of expressing the above-described genes, including substantially similar derivatives thereof.

Numerous vector constructs, including all or part of the nucleotide sequences of a native or derivative *sefA*, *sefB*, *sefC*, *sefD*, *sefU*$_2$, *sefU*$_1$, *agfA*, *tctA*, *tctB*, and/or *tctC* genes, as described above, can be prepared as a matter of convenience. Within the context of the present invention, a DNA construct is understood to refer to a DNA molecule, or a clone of such a molecule (either single-stranded or double-stranded), that has been modified through human intervention to contain segments of DNA combined and juxtaposed in a manner that as a whole would not otherwise exist in nature. Vector constructs of the present invention comprise a first DNA segment encoding one or more of the *sefA*, *sefB*, *sefC*, *sefD*, *sefU*$_1$, *sefU*$_2$, *agfA*, *tctA*, *tctB*, and/or *tctC* genes operably linked to additional DNA segments required for the expression of the first DNA segment. Within the context of the present invention, additional DNA segments will include a promoter and will generally include transcription terminators, and may further include enhancers and other elements.

Mutations in nucleotide sequences constructed for expression of variant proteins preferably preserve the reading frame of the coding sequences. Furthermore, the mutations will preferably not create complementary regions that could hybridize to

produce secondary mRNA structures, such as loops or hairpins, that would adversely affect translation of the mRNA. Although a mutation site may be predetermined, it is not necessary that the nature of the mutation *per se* be predetermined. For example, in order to select for optimum characteristics of mutants at a given site, random mutagenesis may be conducted at the target codon and the expressed mutants screened for indicative biological activity.

Mutations may be introduced at particular loci by synthesizing oligonucleotides containing a mutant sequence, flanked by restriction sites enabling ligation to fragments of the native sequence. Following ligation, the resulting reconstructed sequence encodes a derivative having the desired amino acid insertion, substitution, or deletion.

Alternatively, oligonucleotide-directed site-specific mutagenesis procedures may be employed to provide an altered gene having particular codons altered according to the substitution, deletion, or insertion required. Exemplary methods of making the alterations set forth above are disclosed by Walder et al. (*Gene 42*:133, 1986); Bauer et al. (*Gene 37*:73, 1985); Craik (*BioTechniques*, January 1985, 12-19); Smith et al. (*Genetic Engineering: Principles and Methods*, Plenum Press, 1981); and Sambrook et al. (*supra*).

The primary amino acid structure of the above described proteins may also be modified by forming covalent or aggregative conjugates with other chemical moieties, such as glycosyl groups, lipids, phosphate, acetyl groups, or with other proteins or polypeptides.

Within a further embodiment, the above described proteins may be fused with other peptides that facilitate purification or identification of these proteins. For example, a protein can be prepared as a fusion protein with the FLAG polypeptide sequence (*see* U.S. Patent No. 4,851,341; *see also* Hopp et al., *Bio/Technology 6*:1204, 1988). The FLAG polypeptide sequence is highly antigenic and provides an epitope for binding by a specific monoclonal antibody, enabling rapid purification of the expressed recombinant protein. This sequence is also specifically cleaved by bovine mucosal enterokinase at the residue immediately following the Asp-Lys pairing.

B. Expression Vectors

One type of vector construct, known as an expression vector, can contain DNA segments necessary to direct the secretion of a polypeptide of interest. Such DNA segments can include at least one secretory signal sequence. Preferred secretory signals include the yeast alpha factor signal sequence (pre-pro sequence; Kurjan and Herskowitz, *Cell 30*:933-943, 1982; Kurjan et al., U.S. Patent No. 4,546,082; Brake,

EP 116,201), the Pho5 signal sequence (Beck et al., WO 86/00637), the Suc2 signal sequence (Carlson et al., *Mol. Cell. Biol. 3*:439-447, 1983), the α-2 plasmin inhibitor signal sequence (Tone et al., *J. Biochem. (Tokyo) 102*:1033-1042, 1987), the tissue plasminogen activator signal sequence (Pennica et al., *Nature 301*:214-221, 1983), the *E. coli* PhoA signal sequence (Yuan et al., *J. Biol. Chem. 265*:13528-13552, 1990), or any of the other bacterial signal sequences known in the art, such as those reviewed by Oliver (*Ann. Rev. Microbiol. 39*:615-649, 1985). Alternatively, a secretory signal sequence can be synthesized according to the rules established, for example, by von Heinje (*Eur. J. Biochem. 133*:17-21, 1983; *J. Mol. Biol. 184*:99-105, 1985; *Nuc. Acids Res. 14*:4683-4690, 1986). Secretory signal sequences can be used singly or in combination.

For expression, a DNA molecule as described above is inserted into a suitable vector construct, which in turn is used to transform or transfect appropriate host cells for expression. Host cells suitable for use in practicing the present invention include mammalian, avian, plant, insect, bacterial and fungal cells. Preferred eukaryotic cells include cultured mammalian cell lines (*e.g.*, rodent or human cell lines) and fungal cells, including species of yeast (*e.g.*, *Saccharomyces* spp., particularly *S. cerevisiae*, *Schizosaccharomyces* spp., or *Kluyveromyces* spp.) or filamentous fungi (*e.g.*, *Aspergillus* spp., *Neurospora* spp.). Strains of the yeast *Saccharomyces cerevisiae* are particularly preferred. Methods for producing recombinant proteins in a variety of prokaryotic and eukaryotic host cells are generally known in the art (*see*, "Gene Expression Technology," *Methods in Enzymology*, Vol. 185, Goeddel (ed.), Academic Press, San Diego, Calif., 1990; *see also*, "Guide to Yeast Genetics and Molecular Biology," *Methods in Enzymology*, Guthrie and Fink (eds.), Academic Press, San Diego, Calif., 1991). In general, a host cell will be selected on the basis of its ability to produce the protein of interest at a high level or its ability to carry out at least some of the processing steps necessary for the biological activity of the protein. In this way, the number of cloned DNA sequences that must be introduced into the host cell can be minimized and overall yield of biologically active protein can be maximized.

Suitable yeast vectors for use in the present invention include YRp7 (Struhl et al., *Proc. Natl. Acad. Sci. USA 76*:1035-1039, 1978), YEp13 (Broach et al., *Gene 8*:121-133, 1979), pJDB249 and pJDB219 (Beggs, *Nature 275*:104-108, 1978) and derivatives thereof. Such vectors will generally include a selectable marker, which may be one of any number of genes that exhibit a dominant phenotype for which a phenotypic assay exists to enable transformants to be selected. Preferred selectable markers are those that complement host cell aux trophy, provide antibiotic resistance or enable a cell to utilize specific carbon sources, and include *leu2* (Broach et al., *ibid.*),

ura3 (Botstein et al., *Gene 8*:17, 1979), or *his3* (Struhl et al., *ibid.*). Another suitable selectable marker is the *cat* gene, which confers chloramphenicol resistance on yeast cells.

Preferred promoters for use in yeast include promoters from yeast glycolytic genes (Hitzeman et al., *J. Biol. Chem. 255*:12073-12080, 1980; Alber and Kawasaki, *J. Mol. Appl. Genet. 1*:419-434, 1982) or alcohol dehydrogenase genes (Young et al., in *Genetic Engineering of Microorganisms for Chemicals*, Hollaender et al. (eds.), p. 355, Plenum, New York, 1982; Ammerer, *Meth. Enzymol. 101*:192-201, 1983). The expression units may also include a transcriptional terminator.

In addition to yeast, proteins of the present invention can be expressed in filamentous fungi, for example, strains of the fungi *Aspergillus* (McKnight et al., U.S. Patent No. 4,935,349, which is incorporated herein by reference). Examples of useful promoters include those derived from *Aspergillus nidulans* glycolytic genes, such as the *adh3* promoter (McKnight et al., *EMBO J. 4*:2093-2099, 1985). An example of a suitable terminator is the *adh3* terminator (McKnight et al., *ibid.*, 1985). The expression units utilizing such components are cloned into vectors that are capable of insertion into the chromosomal DNA of *Aspergillus*.

Techniques for transforming fungi are well known in the literature, and have been described, for instance, by Beggs (*ibid.*), Hinnen et al. (*Proc. Natl. Acad. Sci. USA 75*:1929-1933, 1978), Yelton et al. (*Proc. Natl. Acad. Sci. USA 81*:1740-1747, 1984), and Russell (*Nature 301*:167-169, 1983). The genotype of the host cell will generally contain a genetic defect that is complemented by the selectable marker present on the expression vector. Choice of a particular host and selectable marker is well within the level of ordinary skill in the art.

In addition to fungal cells, cultured mammalian cells may be used as host cells within the present invention. Preferred cultured mammalian cells for use in the present invention include the COS-1 (ATCC No. CRL 1650), COS-7 (ATCC No. CRL 1651), BHK (ATCC No. CRL 1632), and 293 (ATCC No. CRL 1573; Graham et al., *J. Gen. Virol. 36*:59-72, 1977) cell lines. A preferred BHK cell line is the BHK 570 cell line (deposited with the American Type Culture Collection under accession number CRL 10314). In addition, a number of other mammalian cell lines may be used within the present invention, including Rat Hep I (ATCC No. CRL 1600), Rat Hep II (ATCC No. CRL 1548), TCMK (ATCC No. CCL 139), Human lung (ATCC No. CCL 75.1), Human hepatoma (ATCC No. HTB-52), Hep G2 (ATCC No. HB 8065), Mouse liver (ATCC No. CCL 29.1), NCTC 1469 (ATCC No. CCL 9.1), SP2/0-Ag14 (ATCC No. 1581), HIT-T15 (ATCC No. CRL 1777), and RINm 5AHT$_2$B (Orskov and Nielson, *FEBS 229*(1):175-178, 1988).

Mammalian expression vectors for use in carrying out the present invention will include a promoter capable of directing the transcription of a cloned gene or cDNA. Preferred promoters include viral promoters and cellular promoters. Viral promoters include the immediate early cytomegalovirus promoter (Boshart et al., *Cell 41*:521-530, 1985) and the SV40 promoter (Subramani et al., *Mol. Cell. Biol. 1*:854-864, 1981). Cellular promoters include the mouse metallothionein-1 promoter (Palmiter et al., U.S. Patent No. 4,579,821), a mouse V_j promoter (Bergman et al., *Proc. Natl. Acad. Sci. USA 81*:7041-7045, 1983; Grant et al., *Nuc. Acids Res. 15*:5496, 1987) and a mouse V_H promoter (Loh et al., *Cell 33*:85-93, 1983). Such expression vectors may also contain a set of RNA splice sites located downstream from the promoter and upstream from the DNA sequence encoding the peptide or protein of interest. Preferred RNA splice sites may be obtained from adenovirus and/or immunoglobulin genes. Also contained in the expression vectors is a polyadenylation signal located downstream of the coding sequence of interest. Suitable polyadenylation signals include the early or late polyadenylation signals from SV40 (Kaufman and Sharp, *ibid.*), the polyadenylation signal from the Adenovirus 5 E1B region and the human growth hormone gene terminator (DeNoto et al., *Nuc. Acids Res. 9*:3719-3730, 1981). The expression vectors may include a noncoding viral leader sequence, such as the Adenovirus 2 tripartite leader, located between the promoter and the RNA splice sites. Preferred vectors may also include enhancer sequences, such as the SV40 enhancer and the mouse l enhancer (Gillies, *Cell 33*:717-728, 1983). Expression vectors may also include sequences encoding the adenovirus VA RNAs. Suitable expression vectors can be obtained from commercial sources (*e.g.*, Stratagene, La Jolla, Calif.).

Vector constructs comprising cloned DNA sequences can be introduced into cultured mammalian cells by, for example, calcium phosphate-mediated transfection (Wigler et al., *Cell 14*:725, 1978; Corsaro and Pearson, *Somatic Cell Genetics 7*:603, 1981; Graham and Van der Eb, *Virology 52*:456, 1973), electroporation (Neumann et al., *EMBO J. 1*:841-845, 1982), or DEAE-dextran mediated transfection (Ausubel et al. (eds.), *Current Protocols in Molecular Biology*, John Wiley and Sons, Inc., NY, 1987), which are incorporated herein by reference. To identify cells that have stably integrated the cloned DNA, a selectable marker is generally introduced into the cells along with the gene or cDNA of interest. Preferred selectable markers for use in cultured mammalian cells include genes that confer resistance to drugs, such as neomycin, hygromycin, and methotrexate. The selectable marker may be an amplifiable selectable marker. Preferred amplifiable selectable markers are the DHFR gene and the neomycin resistance gene. Selectable markers are reviewed by Thilly (*Mammalian Cell Technology*, Butterworth Publishers, Stoneham, MA, which is incorp rated herein by

reference). The choice of selectable markers is well within the level of ordinary skill in the art.

As discussed further below, naked vector constructs can also be taken up by muscular cells subsequent to injection into the muscle of a mammal (or other animals).

Selectable markers may be introduced into the cell on a separate vector at the same time as the *sefA*, *B*, *sefC*, *sefD*, *agfA*, *tctA*, *tctB*, and/or *tctC* genes sequences, or they may be introduced on the same vector. If on the same vector, the selectable marker and the *sefA*, *sefB*, *sefC*, *sefD*, *agfA*, *tctA*, *tctB*, and/or *tctC* genes sequences may be under the control of different promoters or the same promoter, the latter arrangement producing a dicistronic message. Constructs of this type are known in the art (for example, Levinson and Simonsen, U.S. Patent No. 4,713,339). It can also be advantageous to add additional DNA, known as "carrier DNA," to the mixture that is introduced into the cells.

Mammalian cells containing a suitable vector are allowed to grow for a period of time, typically 1-2 days, to begin expressing the DNA sequence(s) of interest. Drug selection is then applied to select for growth of cells that are expressing the selectable marker in a stable fashion. For cells that have been transfected with an amplifiable, selectable marker the drug concentration may be increased in a stepwise manner to select for increased copy number of the cloned sequences, thereby increasing expression levels. Cells expressing the introduced sequences are selected and screened for production of the protein of interest in the desired form or at the desired level. Cells that satisfy these criteria can then be cloned and scaled up for production.

The preferred prokaryotic host cell for use in expressing the gene sequences of the present invention is *Salmonella*. Other preferred host cells include strains of the bacteria *E. coli*, although *Bacillus*, *Shigella* and other genera are also useful. Techniques for transforming these hosts and expressing foreign DNA sequences cloned therein are well known in the art (*see, e.g.*, Maniatis et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1982, which is incorporated herein by reference; or Sambrook et al., *supra*). Vectors used for expressing cloned DNA sequences in bacterial hosts will generally contain a selectable marker, such as a gene for antibiotic resistance, and a promoter that functions in the host cell. Appropriate promoters include the *trp* (Nichols and Yanofsky, *Meth. Enzymol. 101*:155-164, 1983), *lac* (Casadaban et al., *J. Bacteriol. 143*:971-980, 1980), and phage λ (Queen, *J. Mol. Appl. Genet. 2*:1-10, 1983) promoter systems. Plasmids useful for transforming bacteria include the pUC plasmids (Messing, *Meth. Enzymol. 101*:20-78, 1983; Vieira and Messing, *Gene 19*:259-268, 1982), pBR322 (Bolivar et al., *Gene 2*:95-

113, 1977), pCQV2 (Queen, *ibid.*), and derivatives thereof. Plasmids may contain both viral and bacterial elements.

Given the teachings provided herein, promoters, terminators and methods for introducing expression vectors encoding *sefA, B, sefC, sefD, agfA, tctA, tctB*, and/or *tctC* genes of the present invention into plant, avian, fish and insect cells would be evident to those of skill in the art. The use of baculoviruses, for example, as vectors for expressing heterologous DNA sequences in insect cells has been reviewed by Atkinson et al. (*Pestic. Sci. 28*:215-224,1990). In addition, the use of *Agrobacterium rhizogenes* as vectors for expressing genes in plant cells has been reviewed by Sinkar et al. (*J. Biosci. (Bangalore) 11*:47-58, 1987).

Host cells containing vector constructs of the present invention are then cultured to express a DNA molecule as described above. The cells are cultured according to standard methods in a culture medium containing nutrients required for growth of the chosen host cells. A variety of suitable media are known in the art and generally include a carbon source, a nitrogen source, essential amino acids, vitamins and minerals, as well as other components, *e.g.*, growth factors or serum, that may be required by the particular host cells. The growth medium will generally select for cells containing the DNA construct(s) by, for example, drug selection or deficiency in an essential nutrient which is complemented by the selectable marker on the DNA construct or co-transfected with the DNA construct.

Suitable growth conditions for yeast cells, for example, include culturing in a chemically defined medium, comprising a nitrogen source, which may be a non-amino acid nitrogen source or a yeast extract, inorganic salts, vitamins and essential amino acid supplements at a temperature between 4°C and 37°C, with 30°C being particularly preferred. The pH of the medium is preferably maintained at a pH greater than 2 and less than 8, more preferably pH 5-6. Methods for maintaining a stable pH include buffering and constant pH control. Preferred agents for pH control include sodium hydroxide. Preferred buffering agents include succinic acid and Bis-Tris (Sigma Chemical Co., St. Louis, Mo.). Due to the tendency of yeast host cells to hyperglycosylate heterologous proteins, it may be preferable to express the *sefA, B, sefC, sefD, agfA, tctA, tctB*, and/or *tctC* genes of the present invention in yeast cells having a defect in a gene required for asparagine-linked glycosylation. Such cells are preferably grown in a medium containing an osmotic stabilizer. A preferred osmotic stabilizer is sorbitol supplemented into the medium at a concentration between 0.1 M and 1.5 M, preferably at 0.5 M or 1.0 M.

Cultured mammalian cells are generally cultured in commercially available serum-containing or serum-free media. Selection of a medium and growth

conditions appropriate for the particular cell line used is well within the level of ordinary skill in the art.

III. Proteins

A. Proteins Generally

As noted above, the present invention also provides isolated proteins. Within the context of the present invention, such proteins are understood to include the whole, or portions, of a gene product derived from one or more of the *sefA*, *sefB*, *sefC*, *sefD*, *agfA*, *tctA*, *tctB*, and/or *tctC* genes, or derivatives thereof as discussed above. Where the protein is a portion of a native gene or is encoded by derivative of a native gene, the protein maintains substantially the same biological activity of the native protein. The structure of the proteins corresponding to the *sefA*, *sefB*, *sefC*, *sefD*, *agfA*, *tctA*, *tctB*, and/or *tctC* genes can be predicted from the primary translation products using the hydrophobicity plot function of, for example, P/C Gene or Intelligenetics Suite (Intelligenetics, Mountain View, Calif.), or according to the methods described by Kyte and Doolittle (*J. Mol. Biol. 157*:105-132, 1982).

B. Purification of Proteins

Proteins can be isolated by, among other methods, culturing suitable host and vector systems to produce the recombinant translation products of the present invention. Supernatants from such cell lines, or protein inclusions or whole cells where the protein is not excreted into the supernatant, can then be treated by a variety of purification procedures in order to isolate the desired proteins. For example, the supernatant may be first concentrated using commercially available protein concentration filters, such as an Amicon or Millipore Pellicon ultrafiltration unit. Following concentration, the concentrate may be applied to a suitable purification matrix such as, for example, an anti-protein antibody bound to a suitable support. Alternatively, anion or cation exchange resins may be employed in order to purify the protein. As a further alternative, one or more reverse-phase high performance liquid chromatography (RP-HPLC) steps may be employed to further purify the protein. Other methods of isolating the proteins of the present invention are well known in the skill of the art.

A protein is deemed to be "isolated" within the context of the present invention if no other (undesired) protein is detected pursuant to SDS-PAGE analysis followed by coomassie blue staining. Within other embodiments, the desired protein can

be isolated such that no other (undesired) protein or LPS is detected pursuant to SDS-PAGE analysis followed by silver staining.

IV. Compositions Capable of Eliciting an Immune Response, and Vaccines

A. Peptide Compositions

The present invention provides isolated compositions capable of eliciting an immune response, such as immunogens, comprising the amino acid sequences shown in Figures 2 to 7B, or substantial equivalents thereof. Various portions of such amino acid sequences can also be utilized within the context of the present invention. The immunogen may be selected from a portion of the amino acid sequences that is as small as 5 amino acids or as large as about 40 to 50 amino acids, but preferably the portion will be about 12 to 35 amino acids in length. In preferred embodiments, the immunogen comprises a GVVPQ amino acid sequence or a *sefD* amino acid sequence.

As will be understood by one of ordinary skill in the art, slight deviations of the amino acid sequences can be made without affecting the immunogenicity of the immunogen. Substantial equivalents of the above proteins include conservative substitutions of amino acids that maintain substantially the same charge and hydrophobicity as the original amino acid. Conservative substitutions include replacement of valine for isoleucine or leucine, and aspartic acid for glutamic acid, as well as other substitutions of a similar nature (*See* Dayhoff et al. (ed.), "Atlas of Protein Sequence and Structure," *Natl. Biomed. Res. Fdn.*, 1978).

Fimbriae, preferably one or more of SEF14, SEF17, SEF18 and SEF21, when expressed on whole cell (either viable whole cells or in bacterin form) or when presented in a purified form, are capable of eliciting an immune response in animals. In addition, the proteins can be used to generate either polyclonal or monoclonal antibodies. In a prefered embodiment, the isotype of the antibodies (preferably monoclonal antibodies) is IgA. Thus, the fimbriae can be an important bacterial cell component in generating host mucosal immunity to *Salmonella*, an enteric pathogen.

As will be evident to one of ordinary skill in the art, the immunogens listed above, including their substantial equivalents, may stimulate different levels of response in different animals. The immunogens listed above, including their substantial equivalents, can be tested for effectiveness as a vaccine in experiments as described below in the Examples. These experiments include the T-cell proliferation assays, determination of lymphokine production after stimulation, and immunoprotection trials. Briefly, T-cell proliferation assays can be utilized as an indicator of potential for cell mediated immunity. Additionally, evidence of lymphokine production after stimulation

by an immunogen can be utilized to determine the potential for protection provided by an immunogen.

Finally, as described below, trials can be performed in order to determine the level of elicitation of an immune response, including actual immunoprotection, in animals. In the case of humans however, instead of initial immunoprotection trials it is preferred to first screen peripheral blood lymphocytes (PBLs) from patients infected with *Salmonella* in the following manner. Briefly, PBLs can be isolated from diluted whole blood using Ficoll density gradient centrifugation and utilized in cell proliferation studies with ^{3}H-thymidine as described below. Positive peptides are then selected and utilized in primate trials.

The immunogens or peptides of the present invention can be readily produced utilizing many techniques well known in the art (*see* Sambrook et al., *supra*, *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory Press, 1989). Particularly preferred is the synthesis of the immunogens utilizing conventional peptide synthesizers.

B. Attenuated Bacterial Compositions

1. Compostitions using attenuated *Salmonella*

As discussed above, an aspect of the present invention is the induction of mutations in the *tctI*, *tctII* or *tctIII* operons, which are responsible for citrate transport in *Salmonella*. Because *Salmonella* containing such mutations are not able to scavenge for certain nutrients from a host cell, the *Salmonella* is typically unable to proliferate within a host cell. However, a *Salmonella* having an inoperative tricarboxylic acid transport system is still capable of being taken into a host cell, giving an infected cell, and is able to persist within the host cell for some time, thereby allowing the *Salmonella* to elicit a response from the immune system of the host animal. Such a *Salmonella* is known as an attenuated *Salmonella*.

In an alternative embodiment, the *Salmonella* is attenuated via introducing a mutation into two or more of its fimbrial genes, *sefA*, *sefD*, *agfA*, and *fimA*, wherein the mutation effectively prevents production of the selected fimbriae by the *Salmonella*. In preferred embodiments, the mutations are induced in the *fimA* gene (*SEF21*) and the *agfA* gene (SEF17). In an alternative preferred embodiment, the mutations are induced in the *fimA* (SEF21), *agfA* gene (SEF17) and *sefD* gene (SEF18). Further, where *S. enteritidis* is a concern, the group of mutations will preferably include a mutation in the *sefA* gene (SEF14). Alternatively, because SefA protein and SEF14 fimbriae are known to induce a T-cell response, the group of mutations may also include

a mutation in *sefA* that permits production of SEF14 but disrupts any toxic effect that may be due to SEF14. Preferably, such a mutation in *sefA* is caused by inserting a foreign nucleic acid molecule in an appropriate portion of the *sefA* gene.

Other mutations that provide an attenuated *Salmonella* are also known and, while not prefered, can be useful within some aspects of the present invention. Examples of such attenuating mutations include the following: *galE* mutants, which lack UDP galactose epimerase (such mutants also reversibly lack LPS and the Vi antigen) (Nnalue and Stocker, *Microb. Path. 7*:299-310, 1989; Germanier and Rurer, *J. Infect. Dis. 131*:553-558, 1975); *aroA* or Δ*aroA* mutants, which are blocked in aromatic amino acid biosynthesis, as well as other *aro* genes including *aroC* and *aroD* (Hoiseth and Stocker, *Nature 191*:238-239, 1981; Dougan et al., *Molec. Gen. Genet. 207*:402-405, 1987; Brown et al., *J. Infect. Dis. 155*:86-92, 1987); Jones et al., *Vaccine 9*:29-34, 1991; Mukkur et al., *J. Med. Microbiol. 34*:57-62, 1991; Robertsson et al., *Infect. Immun. 41*:742-750, 1983; Cooper et al., *Microb. Pathog. 9*:255-265, 1990); Δ*cya*Δ*crp* combined deletion mutants, which lack adenylate cyclase and cAMP receptor protein (Curtiss and Kelly, *Infect. Immun. 5*:3035, 1987); Δ*phoP* mutants, which lack the regulatory gene involved in acid phosphatase production: (Galan and Curtiss. *Microb. Path.* 6:433, 1989; Miller et al. *Res. Microbiol.* 141:817, 1990); *purA* or Δ*pur* purine biosynthesis mutations, which are typically not useful alone, but rather when used in combination with *aroA* and *galE* (Nnalue and Stocker, *Infect. Immun. 55*:955, 1987; Sigward et al., *Infect. Immun. 57*:1858-1861, 1989); Δcdt mutants, which have a mutation of the chromosome that blocks *Salmonella* from colonizing deep tissue (Curtiss et al., *Vet. Microbiol. 37*:397-405, 1993); *phoP*^{pc} mutants, which are constitutive producers of *phoP* (Miller and Mekalanos, *J. Bacteriol. 172*:2485-2490, 1990.); streptomycin independent reverse mutants (Pardon, *Res. Microbiol. 141*:945-953, 1990); attachment/invasion deficient mutants of *S. choleraesuis* (Wilson et al., *Res. Microbiol. 141*:827-830, 1990); temperature-sensitive *S. enteritidis* mutants (Onozuka et al., *Int J. Immunopharm. 11*:781-787, 1989); *his* mutants in combination with Δ*aroA* and Δ*pur* mutants (Edwards and Stocker., *J. Bacteriol. 170*:3991-3997, 1989); Asp$^-$, Hst, Rbt, Rtt (asparagine auxotroph, high sensitivity to tensides, reversion to bile tolerance, reversion to tenside tolerance) mutants, which have various mutations that are useful to generate strains with graded attenuation: (Linde et al., *Vaccine 8*:278-282, 1990); *htrA* mutants, which lack a stress protein serine protease: (Strahan et al., *Microb. Path. 12*:311-317, 1992); and, mutants that lack the virulence plasmid (Nakamura et al., *Infect. Immun.* 50:586, 1985; Barrow, *Infect. Immun.* 58:2283, 1990)

In preferred embodiments, the host animal is a fish or a warm-blooded animal, and further preferably is a food animal such as poultry, swine or cattle, or even further preferably a human being.

In a preferred embodiment, the attenuated *Salmonella* have been engineered to express a foreign gene (including portions thereof) fused to one or more genes that code for the fimbrin proteins SefA, SefD or AgfA. In a preferred embodiment, the fusion protein comprises a portion of the SEF17 fimbrin, which is encoded by the *agfA* gene operably linked in open reading frame to the gene for the foreign antigen. In an alternative preferred embodiment, the fusion protein comprises a portion of the SEF14 fimbrin, encoded by the *sefA* gene, the SEF18 fimbrin, encoded by the *sefD* gene, or the SEF21 fimbrin, encoded by the *fimA* gene. By "foreign antigen" it is meant an antigen foreign to the host cell and not typically expressed in a *Salmonella* fimbriae. Preferably, the foreign antigen is foreign to both the host cell and the *Salmonella*. The "foreign antigen" includes a whole antigen, as well as a portion of an antigen, comprising numerous epitopes, and also includes an antigen, or portion of an antigen, that is a single epitope.

In a preferred embodiment, the attenuated *Salmonella* composition, including an attenuated *Salmonella* composition comprising a foreign antigen, is an LPS O-polysaccharide deficient strain of *Salmonella* able to express SEF18 or SEF17 fimbriae (which include the foreign antigen, where such is present). In alternative embodiments, the fusion proteins are not expressed as fimbriae, but rather as protein aggregates, inclusion bodies or stable polymers of AgfA, SefA, SefD, SefC, SefB, TctC, TctB or TctA. In a further alternative embodiment, the fusion protein comprises more than one foreign antigen.

2. COMPOSITIONS USING ATTENUATED *E. COLI*

In an alternative embodiment, the attenuated bacteria can be an attenuated *E. coli* able to express one or more *Salmonella sefA, sefD, sefC, sefB, agfA, tctC, tctB* or *tctA* fimbriae genes. In a preferred embodiment, the expressed gene is the *agfA* gene. In alternative preferred embodiments, the expressed gene is the *sefA gene*, the *sefD* gene, or the *fimA* gene. In an alternative embodiment, the peptide corresponding to the desired gene is not displayed as an intact fimbriae, but is still able to elicit an immune response from the host organism. In preferred embodiments, the host animal is a fish or a warm-blooded animal, and further preferably is a food animal such as poultry, swine or cattle. Further preferably, the host animal is a human being.

In a preferred embodiment, the *E. coli* or other bacterial host cell expresses a *Salmonella* AgfA, FimA, SefA, SefD, SefC, SefB, TctC, TctB or TctA

protein fused to one or more foreign antigens. The *Salmonella* protein is further preferably a fimbrin protein, further preferably SefA, AgfA, SefD or FimA. Further, preferably both the *Salmonella*-based antigen and the foreign antigen are able to elicit a response from the immune system of the host animal, yielding a multipurpose composition/immunogen.

C. Nucleic Acid Vaccines

1. Generally

Direct injection, or other appropriate introduction, of one or more of the *sefU*$_2$, *sefU*$_1$, *sefA*, *sefB*, *sefC*, *sefD* *tctC*, *tctB*, *tctA*, or *agfA* genes into an animal can elicit an immune response in the animal, and preferably vaccinate, against the peptide that is expressed from the given gene, and therefore *Salmonella*. In one embodiment, the nucleic acid that is injected further comprises an antigen from a foreign microbe, thereby providing a composition able to elicit an immune response against microbes in addition to *Salmonella*.

In an example of this procedure, naked DNA is introduced into an appropriate cell, such as a muscle cell, where it produces protein that is then displayed on the surface of the cell, thereby eliciting a response from host cytotoxic T-lymphocytes (CTLs). This provides an advantage over traditional immunogens wherein the elicited response comprises specific antibodies. Specific antibodies are generally strain-specific and cannot recognize the corresponding antigen on a different strain. CTLs, on the other hand, are specific for conserved antigens and can respond to different strains expressing a corresponding antigen ("Heterologous protection against influenza by injection of DNA encoding a viral protein," Ulmer et al., *Science 259*:1745-1749 (1993); "Expression of recombinant genes in myocardium *in vivo* after direct injection of DNA," Lin et al., *Circulation 82*:2217-21 (1990); "Long-term persistence of plasma DNA and foreign gene expression in mouse muscle," Wolff et al., *Human Mol. Gen. 1*:363-69 (1992)).

In preferred embodiments, the vector construct containing one or more of the gene sequences of the present invention is naked and is injected into a muscle of an animal, resulting in the uptake of the vector construct by the muscle cells of the animal, and expression of the protein encoded by the DNA. In preferred embodiments, the vector construct is injected into fish or warm-blooded animals, including birds, mice, rats, primates and human beings. Further, it is preferred that the vector construct is a DNA expression vector wherein the desired gene is under the control of a Rous sarcoma virus (RSV) or cytomegalo virus (CMV) promoter (Ulmer et al., *supra*; Lin et al.).

Further preferably, the plasmid resists incorporation into the chromosomal DNA of the animal's cell, and does not replicate within the animal's cell (Wolff et al., *supra*).

Upon introduction of the naked vector construct into the animal's cell, the construct is then able to express the gene that it carries, which gene preferably comprises one (or more) of the *sefU*$_1$, *sefU*$_2$, *sefA*, *sef*B, *sef*C, *sefD*, *tctC*, *tctB*, *tctA*, or *agf*A genes. Accordingly, upon expression of the desired peptide, an immune response is elicited from the host animal. Preferably, the immune response includes $CD8^+$ CTLs able to respond to different strains that exhibit a form of the desired peptide.

In another preferred embodiment, one of the desired proteins discussed above is operably fused in an open reading frame to a second nucleotide sequence (or more) that codes for a foreign antigen, to yield a single dicistronic protein exhibiting antigenic characteristics of both *Salmonella* and another, foreign organism.

D. Administration of Compositions Able to Elicit an Immune Response

1. Peptide Compositions

The present invention provides methods for simulating an immune response in warm-blooded animals comprising administering an effective amount of a pharmaceutical composition comprising an immunogen and a physiologically acceptable carrier or diluent. For purposes of the present invention, warm-blooded animals include, among others, humans, primates, dogs, cats, pigs, sheep, horses, rats, mice, chickens, turkeys and other food animals. The methods can also be used for stimulating an immune response in cold-blooded animals, preferably food animals such as fish.

Many suitable carriers or diluents can be utilized in the present invention, including among others saline, buffered saline, and saline mixed with nonspecific serum albumin. The pharmaceutical composition may also contain other excipient ingredients, including adjuvants, buffers, antioxidants, carbohydrates such as glucose, sucrose, or dextrins, and chelating agents such as EDTA. Within a particularly preferred embodiment, an adjuvant is utilized along with the immunogen. Examples of such adjuvants include alum or aluminum hydroxide for humans.

The amount and frequency of administration can be determined in clinical trials, and will depend upon such factors as the *Salmonella* species against which it is desired to protect, the particular immunogen used, the degree of protection required, and many other factors. In a preferred embodiment, the composition is administered orally, and the attenuated *Salmonella* are taken up by cells, such as cells located in the lumen of the gut. Alternatively, the composition can be parenterally administrated via the subcutaneous route, or via other routes. Depending upon the application, quantities of injected immunogen will vary from 50 μg to several milligrams in an adjuvant vehicle

and preferably about 100 µg to 1 mg, in combination with a physiologically acceptable carrier or diluent. Booster immunizations can be given from 4-6 weeks later.

2. Attenuated Bacterial Compositions

Attenuated bacterial compositions are preferably administered orally in combination with a physiologically acceptable carrier or diluent. However, attenuated bacterial compositions can also be administered via the same mechanisms as peptide-based compositions, such as via injection in combination with an adjuvant.

The following Examples are offered by way of illustration, and not by way of limitation.

EXAMPLES

EXAMPLE 1

Cloning of the *agfA* gene of *Salmonella enteritidis*

An AgfA-negative Tn*phoA* insertion mutant of *S. enteritidis* 27655-3b, named strain 2-7f, was constructed (Collinson et al., "Thin, aggregative fimbriae mediate binding of *Salmonella enteritidis* to fibronectin," *J. Bacteriol. 175*:12-18, 1993). The strain contains an *agfA*-Tn*phoA* gene fusion. Stock cultures of strain 2-7f were prepared using cells from mid-exponential phase cultures that were mixed with 7% glycerol and then stored at -80°C in Luria-Bertani (LB) broth. Isolated colonies of strain 2-7f were obtained by inoculation of solid LB medium followed by incubation at 37°C for 24 hours. An individual colony was inoculated in 2 ml of LB broth contained in a sterile 18 mm x 150 mm test tube and the inoculation was incubated for 48 hours at 37°C under static conditions. For the purpose of isolation of DNA for amplification of *agfA*, cells may be equally effectively prepared by growth in LB broth, Colonization Factor Antigen (CFA) broth, T (tryptone)-medium, other suitable proteolytic digest-based medium, or other medium suitable to support the growth of *Salmonella*. Cells may also be grown under aeration, such as by growth of the culture in an Erlenmeyer or other flask positioned on a rotary or gyratory shaking device. Preferably, the culture is grown at temperatures between 20°C and 37°C.

Cells of strain 2-7f composing the pellicle at the surface of a static culture and cells in suspension were mixed by vortexing for 1 minute. Cells were harvested from 1 ml of this cell suspension (approximately 20 mg wet weight of cells) transferred to a 1.5 ml polypropylene microfuge tube and centrifuged (16,000 x g for 5 min. at 21°

C). The pelleted cells were saved and subsequently resuspended in 1 ml f distilled water. The cells were lysed to release cellular DNA by boiling the 1 ml sample in a sealed microfuge tube for 5 minutes. The cell lysate was partially clarified by centrifugation (16,000 x g for 10 min. at 4°C) in a microfuge to pellet cell debris. The crude DNA preparation (supernatant fraction) was used as a substrate for amplification of a 394 bp DNA fragment encoding the majority of the SEF17 fimbrin subunit, *agf*A.

Standard molecular cloning techniques were performed according to protocols described by Sambrook et al. (*Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press*, Cold Spring Harbor, New York, 1989), or by incorporating minor modifications thereto that are well known in the art. To facilitate cloning of an amplified *agfA* fragment, PCR primers were constructed to include restriction endonuclease cleavage sites. The 5' PCR primer, called TAF1, was a 34 bp biased, degenerate oligonucleotide dGGCGGAAGCTTGAATTCGT[A/C/T]GT[A/C/T]C C[A/G/T]CA[A/G]TGGGG) (SEQ ID No. ___), of which the 17 bases at the 3' end correspond to amino acid residues 2 to 7 of the AgfA N-terminus (the final nucleotide of amino acid 7 is degenerate, and therefore was not made a part of the primer). The amino acid sequence of the N-terminus had been previously determined by Collinson et al. *(supra)*. The underlined sequences were required to create *Hin*dIII and *Eco*RI cleavage sites in the product of DNA amplification. The 3' PCR primer, called TAF2, (dGGGAAAGGTTGAATTCAGGACGCTACTTGTG) (SEQ ID No. ___), into which three nucleotide changes (underlined) were introduced to create an *Eco*RI site in the PCR product, was complementary to the IS*50*$_L$ sequence residing at the junction of Tn*phoA* generated alkaline phosphatase gene fusions. The amplified 394 bp *agfA* fragment was isolated after agarose (1.5%) gel electrophoresis using a 40 mM Tris-acetate, 1 mM EDTA buffer system and then purified using Gene Clean II glassmilk following the 'double Gene Clean' protocol recommended by the manufacturer (Bio 101 Inc., La Jolla, CA).

Approximately 0.5 µg of the amplified *agfA* fragment was cleaved with the restriction endonuclease *Eco*RI at a concentration of 5 units per µg of DNA. In addition, 0.5 µg of the plasmid vector pUC19 (Yannisch-Perron et al., "Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors," *Gene 33*:103-119, 1985) was similarly digested with *Eco*RI. After digestion, the DNA samples were extracted once with 0.75 volume of buffered-phenol and then with 0.75 volume of chloroform to separate enzyme from DNA. The digested DNA fragments were precipitated in the presence of 2.5 volumes of 95% ethanol and 0.3 M sodium acetate, pH 5.4. Residual ethanol was removed from the DNA pellets under vacuum.

The DNA was dissolved in 20 µl of 10 mM TrisHCl - 1 mM EDTA, pH 8.0. Ligation of *agfA* DNA fragments into pUC19 was conducted in 50 µl ligation buffer using 7 units of T4 DNA ligase. Clones containing recombinant plasmids were obtained by transformation of competent cells of *Escherichia coli* strain DH5α obtained from GIBCO BRL Life Technologies Inc. (Burlington, Ontario, Canada) with a sample of ligated DNA. Transformed cells were selected on solid LB medium containing (200 µg/ml) ampicillin, 50 µM IPTG (isopropyl-β-D-thiogalactopyranoside) and 0.005% X-GAL (5-bromo-4-chloro-3-indoyl-β-D-galactoside). Ampicillin resistant colonies that contained the recombinant plasmids were identified by their white color. Recombinant plasmid DNA was purified from transformed cells using a modification of the standard, small scale, alkaline lysis technique described by Sambrook et al. (*supra*). Recombinant plasmids were purified from three individually isolated colonies. These three, apparently identical, plasmids were designated pAGF1, pAGF3 and pAGF4.

In order to obtain a clone containing the full *agfA* gene, chromosomal DNA of *S. enteritidis* strain 27655-3b was purified by CsCl gradient centrifugation and digested separately with *Hin*dIII or *Dra*I and analyzed by Southern hybridization at 65°C using a random-primer, [α-^{32}P]dATP labeled *agfA* PCR fragment according to methods described by Sambrook et al. (*supra*). Following stringent filter washing at 60° C to 62°C as described above, a genomic fragment of approximately 3 kb was identified. *Hin*dIII or *Dra*I digested strain 3b DNA was fractionated according to size by sucrose gradient centrifugation (Sambrook et al., *supra*). DNA fragments contained in the fraction found to hybridize to the [^{32}P]-labeled *agfA* PCR fragment were ligated into M13mpl8 (Yannisch-Perron et al., *supra*) at the *Hin*dIII site or *Sma*I site within the multiple cloning site at 12°C using 12 units of ligase and a total of 1 µg of DNA at a final concentration of 50 µg/µl. Recombinant plaques prepared on a lawn of *E. coli* JM109 (Yannisch-Perron et al., supra), containing *agfA* were identified by dot blot hybridization (Sambrook et al., *supra*) using the [^{32}P]-labeled *agfA* PCR fragment. The recombinant M13mp18 bacteriophage containing *agfA* were plaque purified in the replicative form, double-stranded DNA was purified and the insert fragments encoding *agfA* were cloned into pUC18 (Yannisch-Perron et al., *supra*), and transformed into *E. coli* strain DH5α as described by Sambrook et al. (*supra*). The recombinant plasmid composed of pUC18 and the approximately 3kb *Dra*I fragment of strain 3b DNA was named pDAG6. The recombinant plasmid formed from pUC18 and the approximately 3kb *Hin*dIII fragment of strain 3b genomic DNA was named pHAG10.

EXAMPLE 2

Determination of the DNA sequence of *agfA*

The DNA sequences of both strands of the *agf*A PCR fragment of Example 1 were determined by a modification of the enzymatic, dideoxy-termination sequencing method (Sanger et al., "DNA sequencing with chain-termination inhibitors," *Proc. Natl. Acad. Sci. USA 74*:5463-5467, 1977) using the primers TAF1 and TAF2 and the reagents and protocols supplied in the Sequenase™ Version 2.0 kit (United States Biochemicals, Cleveland, Ohio). The sequence was confirmed by determining the DNA sequence of *agf*A contained in two of the independent plasmid clones. For this determination, an Applied Biosystems model 373A automated DNA sequencer and associated reagents, protocols and software (version 1.10) for thermal cycle sequencing (Applied Biosystems Canada Inc., Mississauga, ON) was used. The universal forward and reverse sequencing primers (Yannisch-Perron et al., *supra*) were used. The DNA sequence of *agfA* encoded on the overlapping *Hin*dIII and *Dra*I fragments cloned in pHAG10 and pDAG6 was determined on both strands using double-strand DNA sequencing techniques.

Computer analysis of the DNA sequence for open reading frames and gene translation was performed using DNA Strider, version 1.1. The start of the open reading frame of the *agfA* gene encoding the mature AgfA fimbrin was recognized by comparison of the translated sequence to the N-terminal amino acid sequence of AgfA. The *agfA* DNA sequence is presented in Figure 7B. The translated DNA sequence of the single open reading frame corresponded precisely to amino acid residues 2 to 31 determined by N-terminal sequencing of AgfA. The amino acid composition of the translated sequence had a similar high glycine content (16%), high combined alanine, serine plus glycine content (37%), low basic amino acid content (4.5%), and nearly 30% asparagine plus aspartic acid content consistent with the total amino acid analysis of native AgfA fimbrin (8). Based on the estimated molecular weight of AgfA of 17 K M_r, and the expected molecular weight of the protein encoded in the 333 bp region of *agfA*, about 12 K - 13 K M_r, it appears that about three quarters of SEF17 fimbrin is represented in the fragment.

EXAMPLE 3

Production and expression of assembled, enzymatically active AgfA-enzyme fusion proteins in *Salmonella enteritidis*

Salmonella enteritidis produce novel thin aggregative fimbriae, known as SEF17, that comprise a class of stable protein polymers requiring treatment with 90% formic acid to depolymerize the fimbriae into their composite fimbrin subunits (Collinson, *supra*). SEF17-negative Tn*phoA* mutants were isolated that were defective in the production of SEF17 fimbriae due to transposon mutagenesis resulting in the fusion of the SEF17 fimbrin gene, *agfA*, with the promoterless alkaline phosphatase gene, *phoA* (Collinson et al., "Thin, aggregative fimbriae mediate binding of *Salmonella enteritidis* to fibronectin," *J. Bacteriol 175*:12-18, 1993). Western blot data indicated that the AgfA-PhoA fusion protein expressed by the Tn*phoA* mutant 2-7f either aggregates or self assembles into polymers because formic acid treatment was required to disassemble these polymers into the AgfA-PhoA subunits.

EXAMPLE 4

Sequencing of *the tctCBA* gene cluster of *Salmonella typhimurium*

The tricarboxylic acid transport (*tctI*) operon of *Salmonella typhimurium* LT2 was isolated on an 8 kb *Eco*RI-*Bam*HI fragment cloned into the vector pBR322 to create the recombinant plasmid pKW101 which conferred a citrate-utilizing (*cit*$^+$) phenotype on *E. coli* MC4100 (Widenhorn et al., "Cloning and promoters of the *Salmonella typhimurium* tricarboxylate transport operon in *Escherichia coli*," *J. Bacteriol. 170*:883-888, 1988). pKW101 was digested with the restriction enzymes *Kpn*I and *Pst*I to yield a fragment of 4.5 kb, containing the entire *tctI* operon, which was inserted into the M13 cloning vector M13mp18 (Yannisch-Perron et al., *supra*) to produce the recombinant clone KS1016. KS1016 was digested with the restriction enzymes *Eco*RI and *Hin*dIII (which cut only in the vector DNA) to release the *tctI* fragment for subcloning into the M13 cloning vector MWB2349 to produce MKS3. (Barnes et al., "Kilo-Sequencing: Creation of an Ordered Nest of Asymmetric Deletions Across a Large Target Sequence Carried on Phage M13," *Meth. Enz. 101*:98-122, 1983.)

To facilitate determining the DNA sequence of *tctCBA*, the 4.5 kb *Eco*RI-*Hin*dIII *tctI* fragment was subcloned from KS1016 into the M13 cl ning vector

MWB2341, whose multiple cloning site is oriented in the opposite direction relative to the binding site of the universal DNA sequencing primer. This resulted in the creation of MKS11.

The recombinant clones MKS3 and MKS11 were subjected to a procedure generating nested deletions (Dale et al., "A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: Application to sequencing the corn mitochondrial 18S rDNA," *Plasmid 13*:31-40, 1985). Briefly, single-stranded recombinant M13 DNA was isolated by standard procedures (Sambrook et al., *supra*). DNA was linearized by *Hin*dIII digestion following hybridization of the oligonucleotide WK-10, 5'-TGAATTAATTCCA CAAGCTTTTTTTTTTT-3'; to MKS3 (SEQ ID No.___) or WK-9, 5'-CGACGGCCAGTGCCAAGCTTTTTTTTTT-3' to MKS11 (SEQ ID No. ___) to create a double stranded restriction endonuclease cleavage site in an otherwise single-stranded molecule.

The linearized DNA was subsequently degraded by the 3' to 5' exonuclease activity of T4 DNA polymerase. Samples were withdrawn at 15, 30, 45, 60, 75, 90 and 105 minutes. The polymerase was inactivated by treating the sample at 65°C for 2 minutes and the samples stored on ice. By withdrawing samples at different times, populations of DNA fragments with differing degrees of deletion were created. All time-point samples were pooled and the DNA was treated with terminal transferase in the presence of ATP to create a polyA tail. The deleted DNA molecules were hybridized with oligonucleotide WK-10 by heating to 65°C for 5 minutes, followed by slow cooling for 30 minutes, then ligated by T4 DNA ligase for 1 hour at 21°C and overnight at 4°C. The ligated fragments were transformed into *E. coli* WB373 made competent by the procedure of Mandel (Mandel and Higa, "Calcium dependent bacteriophage DNA infection," *J. Mol. Biol. 53*:159, 1970). Well-isolated M13 plaques were picked, and grown up overnight (Sambrook et al., *supra*) into 2X-YT broth. The size of each recombinant DNA molecule was estimated by agarose gel electrophoresis. From over 100 plaques analyzed, 35 recombinant phage representing deletions of 0.3 - 4.8 kb were selected for DNA sequence analysis.

DNA sequences were determined using standard modifications of the enzymatic dideoxy termination method of Sanger et al. (*supra*). To resolve the sequence of regions that proved difficult to determine by standard protocols, the sequencing reactions were performed using Sequenase (a chemically modified form of T7 DNA polymerase; US Biochemicals) and either deoxyinosine or 7-deazadeoxyguanosine in place of deoxyguanosine or standard protocol sequencing reactions were analyzed by wedge-gel electrophoresis. Whereas, most sequencing

reactions utilized the universal forward sequencing primer, specific internal primers were also used. The names and sequences of the internal primers used are: WWK-19 5'-GGGCGACTATCGCGTTA-3', WWK-20 5'-AGCCACTTGTAGCGGCC-3', WWK-21 5'-GGAAGTGCATTTTACGT-3', WWK-22 5'-CATGCTGCCAAGACAGG-3', WWK-23 5'-CTTTGGATCTGCCAGGC-3', WWK-24 5'-GCGCCGTCATGATCGCC-3' (SEQ ID Nos. ___, ___, ___, ___, ___, and ___, respectively). The sequences for *tctA*, *tctB* and *tctC* are shown in Figures 4A-4B, 5 and 6A-6B, respectively.

The sequences of *tctA*, *tctB*, and *tctC* were confirmed by automated DNA sequencing using an Applied Biosystems, Inc. Model 373A automated DNA sequencer and the reagents and protocols provided by the manufacturer for cycle-sequencing (Applied Biosystems, Inc., Foster City, CA). DNA oligonucleotide sequencing primers used for this purpose included:

5'TCGGGATGCTGTTCGGCG3' (SEQ. ID. No. _____)
5'CTGCCTGCGGAGTCGGC3' (SEQ. ID. No. _____)
5'GTCGCAAGGCCAAGACCG3' (SEQ. ID. No. _____)
5'GTGTATCGGCACCACCCTG3' (SEQ. ID. No. _____)
5'CCCGGCGATGTTCACCG3' (SEQ. ID. No. _____)
5'CCAATACCGCGCCGGAG3' (SEQ. ID. No. _____)
5'GCGGAGGCAATGATGAGCG3' (SEQ. ID. No. _____)
5'TGCCGCCATACTCACAGCC3' (SEQ. ID. No. _____)
5'TCTTGGCAGCATGATGGCG3' (SEQ. ID. No. _____)
5'CTGGCAATGGTCGCCCG3' (SEQ. ID. No. _____)
5'GCAATCAGCAGCGCAGC3' (SEQ. ID. No. _____)

A restriction map prepared from the DNA sequence appears in Figure 9. The Figure shows the positions of endonuclease cleavage sites for *Apa*I, *Ava*II, *Bgl*II, *Bgl*I, *Bst*II, *Dra*I, *Eco*RV, *Kpn*I, *Nco*I, *Sal*I, *Sma*I, *Sna*BI, *Ssp*I and *Xmn*I relative to the open reading frames of *tctC*, *tctB* and *tctC*.

DNA sequences encoding TctC were recognized by comparison to the complete sequence of TctC protein; the predicted amino acid sequence agreed with the amino acid sequence determined by peptide sequencing studies on the purified protein. The predicted size of the TctB protein based on the nucleotide sequence agreed with the size of an expressed protein as determined by SDS-PAGE.

EXAMPLE 5

Salmonella strains containing mutations in tricarboxylate transport (*tct*) operons

A. Construction of *tctI* mutants in *S. enteritidis*

DNA encoding the *Salmonella typhimurium tctI* operon was cloned into pKW108 ("Expression of the divergent tricarboxylate transport operon (*tctI*) of *Salmonella typhimurium*," *J. Bacteriol. 170*:3223-3227, 1988). The DNA was cleaved with the restriction enzyme *Bgl*II, which cleaves the operon at a single site in the *tctC* gene. The linearized DNA was then digested with exonuclease III and nuclease S1. Samples were withdrawn at selected time points routine in the art and the enzymatic activity was stopped by the addition of EDTA. After all of the samples had been removed, the DNA was pooled, purified, then treated with the Klenow fragment of *E. coli* DNA polymerase I. The DNA was ligated with T4 ligase to give vector constructs that were then transformed into *E. coli* DH5α (Sambrook et al., *supra*).

The resultant transformants were grown up in L-broth supplemented with 100 μg/ml ampicillin overnight and a plasmid preparation was performed by the alkaline lysis procedure (Sambrook et al., *supra*). The population of plasmids was sized on an agarose gel, and shown to consist of molecules of heterogeneous sizes. The plasmid DNA was pooled, then linearized with *Eco*RI and used to electrotransform *S. enteritidis*. The transformants were grown on solid modified Davis minimal medium containing 50 μg/ml fluorocitrate (Fc), to select for transformants that had integrated an inactive *tctI* operon into the cellular chromosome and thus were resistant to the toxic citrate analog fluorocitrate.

Recombinants were screened by colony lifts followed by hybridization to probe complementary to the region that had been deleted. Colonies that hybridized poorly to the probe were picked and grown in L-broth overnight. The *tctC* gene was characterized by PCR using oligonucleotides complementary to the DNA sequence of nucleotides 1459-1481 and 1955-1977 of the *tctI* operon (*see* Figures 6A-6B). One of the transformants tested showed an internal deletion in the *tctC* gene of approximately 300 nucleotides between 1459 and 1977. This bacteria was named *S. enteritidis* strain SL54.

B. Construction of *tctII*$^-$ strain

Constitutive "on" mutants of the *tctII* operon in wild-type *Salmonella typhimurium* SU453 were isolated by plating out cells on Modified Davis Minimal Medium containing 93.4 mM Na^+, with cis-aconitate as the carbon source and no K^+

ions (Na-cac medium). The resultant *tctII*^{on} mutants were named KS823-832. KS823 was used as a recipient strain for P22 *int3* HT12/4 transducing phage (B. Ely et al., "Some improved methods in P22 transduction," *Genetics 76*:625-631, 1974) that had been grown on an *S. typhimurium* strain carrying a *tctI* operon that had been insertionally activated by Tn10, which carries a tetracycline resistance gene (N. Kleckner, "Transposon Tn10 Mobile genetic elements," pp. 261-298, Academic Press, N.Y., 1983). Bacteria that were *tctI*$^{-}$ and *tctII*^{on} were selected for by growth on Na-cac medium in the presence of tetracycline. The tetracycline resistant strains resulting from this cross were called KS838-847.

KS838 was determined to be sensitive to fluorocitrate (Fc). KS838 was grown in the presence of Fc on Na-cac medium, using the disk radial streak assay (Somers et al., "Fluorocitrate-resistant tricarboxylate transport mutants of *Salmonella typhimurium*," *Mol. Gen. Genet. 181*:338-345), and mutants resistant to Fc (Fc^{R}) were collected. These resistant strains were considered *tctII*$^{-}$, and were named KS858-867. Next, the tetracycline resistance of the mobile genetic element Tn10 was removed by making KS858 a recipient strain for P22 transducing phage that had been grown on wild-type *Salmonella* SU453. These phage replace the *tctI*-locus with the wild-type locus. *tctI*$^{+}$ isolates were selected by growing on Davis Modified Minimal Medium supplemented with 20 mM isocitrate as a carbon source. The resultant strains were called KS950-959 and KS970-979, and are phenotypically *tctI*$^{+}$, *tctII*$^{-}$, *tctIII*$^{+}$. These strains contain no foreign DNA or introduced antibiotic resistance genes.

C. Construction of *tctIII*$^{-}$ mutants

Salmonella typhimurium SU453 strains were used, through Tn10 insertion, to produce a series of *tctIII*$^{-}$ isolates named *S. typhimurium* KS1170 and KS1169. The Tn10 elements can be induced to excise through culturing the bacteria on medium containing fusaric acid (S. Maloy et al., "Selection for loss of tetracycline resistance by *Escherichia coli*," *J. Bacteriol. 140*:297-300, 1981). Several strains will suffer a deletion of *tctIII* flanking sequences concurrent with imprecise Tn10 excision and then will be useful as attenuated strains.

EXAMPLE 6

Sequencing and characterization of *sefU$_1$U$_2$ABCD* from the *sef* operon of *Salmonella enteritidis*

A 5.3 kb *Hin*dIII fragment carrying the SEF14 fimbrin structural gene and part of the *sef* operon was subcloned from cos48, a recombinant cosmid carrying 44

kb of *S. enteritidis* chromosomal DNA, into pTZ19R to generate pKX1 (Feutrier et al., "Cloning and expression of a *Salmonella enteritidis* fimbrin gene in *Escherichia coli*," *J. Bacteriol. 170*:4216-4222, 1988; Müller et al., "Fimbriation genes of *Salmonella enteritidis*," *J. Bacteriol. 171*:4648-4654, 1989). A series of overlapping deletion subclones of pKX1 (ΔpKX1) were generated using pKX1 linearized with *Eco*RI and varying degrees of DNAse I digestion according to the method of Lin et al. ("An improved DNA sequencing strategy," *Anal. Biochem. 147*:114-119, 1985), to create a collection of 50 nested deletions. The resulting linear ΔpKX1 subclones were treated with the Klenow fragment of DNA polymerase I and then blunt end ligated with T4 DNA ligase to recircularize the plasmids. The various ΔpKX1 were transformed into *E. coli* XL-1 Blue (Stratagene, La Jolla, CA) using standard procedures (Sambrook et al., *supra*). The ΔpKX1 were purified by standard alkaline lysis procedures (Sambrook et al., *supra*) and run on a 1% agarose gel. A series of ΔpKX1 subclones separated by about 200 to 400 bp in size were chosen and named delA10, delB15, delB23, delC1, delD5, delD8, delD9, delD16, delD19, delE1, and delE21. Large amounts of these plasmids were purified by alkaline lysis for DNA sequencing.

The ΔpKX1 subclones were sequenced by the dideoxynucleotide chain termination method (Sanger et al., "DNA sequencing with chain terminating inhibitors," *Proc. Natl. Acad. Sci. USA 74*:5463-5467, 1977) using T7 DNA polymerase (T7 DNA Polymerase Sequencing Core System, Deaza, Promega, Madison, WI) and deoxyadenosine 5'-[α-^{35}S] triphosphate, (New England Nuclear, Markham, ON) according to the manufacturers' specifications.

The result of DNA sequencing reactions were electrophoresed through a 6% polyacrylamide gel (45 W, 55-60°C) using a discontinuous buffer system (Carninci et al., "A simple discontinuous buffer system for increased resolution and speed in gel electrophoretic analysis of DNA sequence," *Nucleic Acid Res. 18*:204, 1989). Following electrophoresis, the gels were fixed in a solution of 12% methanol and 10% acetic acid for 15 min, and dried onto 3MM paper (Whatman Intl. Ltd., Maidstone, England) under vacuum at 80°C for 2 hours on a Savant gel drying apparatus (Savant Instruments Inc., Farmingdale, NY). Dried gels were exposed to X-Omat K XK-1 film (Kodak, Rochester, NY) and the sequence read directly from the developed films. Both DNA strands were fully sequenced, using the 17 bp reverse primer for the coding strand (GTCATAGCTGTTTCCCG) (Sequence ID No. ___) and 12 custom made internal oligonucleotide primers (ULTRA Diagnostics Corporation, Seattle, WA) for the opposite strand. To complete the sequence for the *sefD* gene, and to sequence the $sefU_1$ and $sefU_2$ genes, a further 10 kb *Kpn*I fragment was obtained from cos48 and

subjected to sequence analysis generally as above. The *sefA, sefB, sefC and sefD* sequences are depicted in Figures 2A-2D.

The genes for $sefU_1$ and $sefU_2$ were similarly sequenced on overlapping subclones and these sequences are depicted in Figures 3A-3B.

The programs contained in MacVector (Intelligenetics, Mountain View, CA) were used to determine the order of the overlapping DNA sequences. DNA Strider™ version 1.1 was used to identify the open reading frames for *sefA*, *sefB* and *sefC*, which were predicted to encode polypeptides of 14,436 M_r, 28,012 M_r and 90,268 M_r, respectively. The predicted amino acid sequences of the SefA, SefB and SefC proteins were compared to proteins listed in the GenBank (release #66.0), SWISS-PROT (release #16.0), and GENPEPT (release #64.3) data bases and the MACAW program (NCBI, Bethesda, MD) to align regions of local similarity among proteins exhibiting similarity.

The adjacent downstream gene, *sefB*, encodes a fimbrial periplasmic chaperone protein.

sefC, the gene immediately downstream of *sefB*, encodes a fimbrial outer membrane protein that contains nine putative membrane-spanning domains.

The nucleotide sequence of DNA immediately downstream of *sefABC* revealed a fourth open reading frame (ORF) designated *sefD*. This gene had the same translational polarity as *sefABC* (Figure 1). In fact, the AUG start codon for *sefD* overlapped the UGA stop codon of *sefC*. The gene organization of the gene cluster, has been confirmed on the chromosome by Southern blot analysis of *KpnI* digested *S. enteriditis* 3b chromosomal DNA hybridized with *sefA* and *sefD* specific probes. Preceding the *sefD* ORF by 8 bp was a consensus Shine-Dalgarno sequence for translation initiation (GGAG). The *sefD* ORF was 447 bp and the predicted molecular weight of the encoded protein, designated SefD, was 16,722 Daltons.

The predicted amino acid sequence of SefD had a putative signal peptidase cleavage site between Ser-24 and Ser-25 as determined by the method of von Heijne (1984). The presence of a putative leader sequence suggested that the protein was exported from the cytoplasm to either the periplasmic space or the outer membrane.

To confirm that *sefABC* encoded proteins of the predicted sizes, proteins were translated in vitro from pKX1. The plasmid-encoded proteins were labeled with 35S-methionine using a cell-free coupled transcription-translation system (Prokaryotic DNA-Directed Transcription-Translation System Kit, Amersham, Oakville, ON) according to the manufacturers' instructions.

Plasmids carrying either the 5.3 kb fragment of the *sef* peron or deletions thereof (delB15, delB23, delD10) were used as DNA templat s (Figure 8B).

Plasmids purified by alkaline lysis (Sambrook et al., *supra*) were incubated with the other reaction components in a final reaction volume of 30 µl and incubated at 37°C for 30 min. Unlabeled methionine (5 µl) was added, the mixture was incubated a further 5 min and then the reaction was terminated by placing the reactions at 0°C. Ten µl of the reaction mixture was added to 2x Laemmli sample buffer (Laemmli, "Cleavage of structural proteins during the assembly of the head of bacteriophage T4," *Nature 227*:680-685, 1970) and then subjected to microdialysis (Marusyk and Sergent, "A simple method of dialysis of small volume samples," *Anal. Biochem. 105*:403-404, 1989) on Millipore filters (VFWP25, Millipore, Millipore Corp., Bedford, MA) for 10 to 15 min prior to SDS-PAGE analysis on 12% gels poured with a 5% stacking gel (Laemmli, *supra*). The acrylamide gel was fixed for 30 min in 7% acetic acid for 30 min, dried onto 3MM paper for 1 hour at 80°C and then exposed to X-Omat-AR5 film (Kodak, Rochester, NY) to visualize the labeled proteins.

Several translation products were identified (Figure 8A, lane 2). The 14K M_r protein was identified on Western blots as SefA (Figure 8A, lane 7). The 90K M_r protein was identified as SefC. The 27K M_r protein was identified as SefB. The 70K, 44K and 40K M_r bands were likely minor degradation products of SefC because these bands were absent when pSC1, which contained a deletion in *sefC*, was used as the template (Figure 8A, lanes 2 and 3). The 16K M_r band seemed to be a minor degradation product of SefB as this band remained when pSC1 was used as a template (Figure 8A, lanes 2, 3). When the three DNase1 deletion subclones, delB15, delB23 and delD10, were each used as templates, the bands for SefB, SefC and their minor degradation products were absent (Figure 8A, lanes 4-6) indicating *sefA* and/or its upstream region is necessary for the expression of *sefB* and *sefC*, as was predicted from the DNA sequence analysis (Figures 1, 2).

To confirm that translation of SefB and SefC was dependent on the presence of *sefA* and/or the region upstream of *sefA*, the transcription start sites for *sefA*, *sefB* and *sefC* were determined. Primer extension studies consistently revealed transcription start sites immediately upstream of *sefA*. These included two major extension products as well as several minor ones. When the primer extension reaction was performed at 50°C, a temperature expected to destabilize secondary structures, reverse transcriptase still stopped at all the sites with the same frequency suggesting that stem-loop structures were not blocking the migration of reverse transcriptase. No transcription start sites could be found immediately upstream of *sefB* or *sefC* (data not shown). These results indicated that the 5' end of the mRNA transcript of *sefABC* was initiated upstream of *sefA*.

EXAMPLE 7

In vivo expression of assembled fimbriae in *E. coli*

Immunogold labeling of *S. enteritidis* cells with polyclonal immune serum raised to purified SEF14 fimbriae revealed thin, filamentous organelles located on the cell surface (Figure 10A). *E. coli* cells hosting pKX1 or cos 48 were shown by Western blot analysis to produce SefA.

To determine whether *E. coli* cells hosting pKX1 or cos48 were assembling intact fimbriae, *E. coli* cells were grown in 2.5 ml of colonization factor antigen static broth culture (Evans et al., "Hemagglutination of human group A erythrocytes by enterotoxigenic *Escherichia coli* isolated from adults with diarrhea: correlation with colonization factor," *Infect. Immun. 18*:330-337, 1977) for 60 h at 37°C. Cells were then mounted on formvar-coated grids, incubated with immune serum to native SEF14 fimbriae, then labeled with protein A-gold (Auroprobe, Pharmacia, Uppsala, Sweden). Labeled cells were then negatively stained with ammonium molybdate.

Immunoelectron microscopic examination of *E. coli* HB101 containing cos48 revealed the presence of immunogold-labeled SEF14 fimbriae on the cells surface (Figure 10B). Similar experiments demonstrated SEF14 fimbriae were produced in *E. coli* containing a cloned 10 kb *Kpn*I fragment encoding *sefU*$_2$*U*$_1$*ABCD*. Examination of *E. coli* JM109 carrying pKX1 revealed the formation of surface blebs that were specifically gold labeled (Figure 10C). *E. coli* carrying 44 kb of *S. enteritidis* DNA (cos48) encompassing the *sefABC* operon displayed intact fimbriae, conversely, *E. coli* carrying the 5.3 kb *Hin*dIII fragment subcloned from cos48 showed that distinguishable SEF14 filamentous fimbriae were not assembled, thus indicating that additional fimbrial genes are required for *in vivo* expression of assembled fimbriae.

EXAMPLE 8

Co-Expression Of The *SefA*, *SefB* and *SefC* Genes

sefB encodes a chaperone protein capable of preventing premature polymerization of fimbrial subunits in the periplasm. *sefB* can be used to chaperone the SefA protein, including the SefA protein fused to a foreign gene product (preferably comprising a foreign antigen) through the inner membrane to the periplasm, where the protein is ushered out of the cell by *sefC*. Such a foreign gene product can also be incorporated in a SEF14-like fimbrial structure.

A 10 kb *Kpn*I fragment from cos48 is cloned into pUC19. The resulting clone is transformed into a competent *E. coli* or *Salmonella* host cell. To demonstrate expression, the host cells are disrupted and subjected to Western blot analysis using antisera to SefA, SefB and SefC proteins. The supernatant (prior to cellular disruption) is also subjected to Western blot analysis using antisera to SefA protein.

To effect transport of protein to the supernatant without requiring the expression of *sefC*, a 1.5 kb *Bsm*I fragment isolated from pTZ19 piIII, is blunted using T4 DNA polymerase and cloned into the *Sma*I site of pUC19. The *Eco*RI/*Hin*dIII fragment of this clone is ligated into the *Eco*RI/*Hin*dIII sites of pINIII11^3-B1 to create a construct where transcription of *sefAsefB* is initiated from the strong *lpp* promoter, which promoter is under control of the *lac* operator.

To demonstrate expression of the genes in *E. coli*, the resulting plasmid is transformed into a "leaky" *E. coli* strain known in the art. Expression is induced from the *lac* operator by the addition of IPTG. Western blot analysis of disrupted cells using antisera to SefA and antisera to SefB shows that both proteins are produced in the *E. coli*.

EXAMPLE 9

Production of Protective Antibodies to *Salmonella enteritidis*

In a preliminary experiment, outbred female mice were immunized with 5×10^5 *S. enteritidis*-3b by interperitoneal injection (IP) on day 0. Blood samples were collected at 10 and 20 days. Western blot analysis indicated that serum antibodies were generated to SefD (SEF18) but no detectable antibodies were found to the other fimbrins.

Whole cells of wild type *S. enteritidis* 3b, an LPS o-polysaccharide deficient strain (3b Tn*phoA*-9) and a SEF17-deficient strain (3b Tn*phoA* 2-7f) were grown under various conditions (T medium, 37°C, 24h; CFA broth, static, 37°C, 48h; LB broth, aerated, 37°C, 24h) to variably express the four fimbrial types SEF14, SEF17, SEF18, and SEF21. Cells were harvested by centrifugation, washed in PBS, resuspended to an OD_{650nm} of 2.0, treated with a final concentration of 3.5% formaldehyde and, incubated overnight at 4°C, the cells were then washed 5 times in PBS, then resuspended to an OD_{650nm} of 1.0. These whole cell bacterins were injected into three week old turkeys using Alhydrogel as an adjuvant. The turkeys were boosted 7 and 14 days later and then bled. The serum was tested for antibodies against all four

fimbrial types on Western blots and it was found that the turkeys made antibodies to SEF14, SEF17 and SEF18.

Purified SEF14, SEF17 and SEF21 were used to raise monoclonal antibodies using standard monoclonal antibody generation techniques (Harlow and Lane, supra). Several monoclonal antibodies were generated to SEF14 (total of 8), SEF17 (total of 2) and SEF 21 (total of 5). One of the monoclonal antibodies generated to SEF14 and two generated to SEF21 were found to be of isotype IgA. This indicates that some fimbriae may specifically illicit a secretory antibody response, which can be advantageous for a vaccine directed toward *Salmonella*, an enteric pathogen.

EXAMPLE 10

Production of polyclonal antibodies to Fimbriae

A standard immunization protocol has been used to raise rabbit polyclonal antibodies to each of the four fimbrial types, namely; SEF14, SEF17, SEF18 and SEF21. Purified, insoluble fimbriae were obtained as previously described (Collinson et al., *supra*), resuspended in phosphate buffered saline and emulsified with Freund's complete adjuvant prior to subcutaneous and intramuscular injection of a 1-month-old female New Zealand White rabbit. The rabbit was subsequently boosted at 1 or 2 week intervals with 50 to 500 µg protein emulsified in Freund's incomplete adjuvant. When the antibody titers were sufficiently high, the serum was collected. These serum antibodies were shown numerous times, via Western blotting techniques, to bind to their respective purified fimbrin subunits, and to bind the native fimbriae on whole cells as ascertained by immunoelecton microscopic observation of protein A gold-labelled fimbriae.

EXAMPLE 11

Creation of a *sefA* (SEF14) Mutant

pDRA, which contains a *Dra*I fragment encompassing the *sefA* gene of *S. enteritidis* inserted in the *Sma*I site of pT7T318, is digested with *Eco*O109I. Two oligonucleotides, LKC3A and LKC3B (Table 1, *infra*) are synthesized, annealed together, phosphorylated and ligated to *Eco*O109I-digested pDRA to inactivate *sefA*. The pDRA::LKC3 construct is digested with *Bgl*II to remove multiple linkers and Sephaglas™ purified following agarose gel electrophoresis. Self-ligation of this construct results in the generation of pKCS2 which contains *sefA* inactivated with a

single linker. pKCS2 is digested with *Eco*RI and *Hinc*II to remove the *sefA*::LKC3 fragment which is then purified prior to electroporation.

To insertionally inactivate *sefA* with a kanamycin (Km) resistance gene, linearized pDra is blunt-ended using the Klenow fragment and ligated to the 1.55 kb *Fsp*I fragment of pKEM containing the kanamycin resistance gene. The resulting plasmid, pKCS1, was digested with *Eco*RI and *Hinc*II to cut out the *sefA*::Km fragment which is purified prior to electroporation.

S enteritidis 3b is transformed with the insertionally inactivated *sefA* gene fragment by electroporation to permit chromosomal *sefA* gene replacement by homologous recombination. Transformants are selected on media containing kanamycin for those transformed with *fimA*::Km or screened for the inability to produce SefA or SEF14 for those transformed with the *fimA*::LKC3. The constructs are confirmed by PCR and DNA sequencing or Western blot analysis.

This method can also be applied using cloned and sequenced chromosomal *sefA* encoding DNA fragments from *S. enteritidis* (or *S. berta, S. dublin, S. gallinarum,* or *S. pullorum).*

EXAMPLE 12

Creation of an *agfA* (SEF17) Mutant

The plasmid pW3, which carries a 394 bp *agfA* fragment, containing a unique *Xcm*I site at bp 156, is digested with *Xcm*I, and the linearized plasmid is purified by agarose gel electrophoresis and Sephaglas™ treatment. Two oligonucleotides, LKC1A and LKC1B (*see* Table 1), synthesized on a Model 391 PCR-Mate™ DNA synthesizer (Applied Biosystems, Mississauga, ON), are mixed and annealed to create a double-stranded DNA fragment, LKC1, with 'sticky-ends' complementary to those generated by *XcmI* digestion. LKC1, which contains a *Bgl*II cleavage site, is phosphorylated using T4 polynucleotide kinase (Pharmacia P-L Biochemicals, Inc., Piscataway, NJ) (Sambrook et al. 1989) and ligated to *Xcm*I-digested pW3 to insertionally inactivate *agfA*. *Bgl*II is used to digest pW3 carrying multiply ligated linkers. Following purification, pW3 carrying the *Bgl*II digested linkers is self-ligated to generate pKCA1. Alternatively, pW3 carrying the *Bgl*II digested linkers was mixed with the 3.4 kb *Hind*III fragment of pHP45ΩCm (carrying a chloramphenicol resistance gene) to generate pKCA2.

The resulting insertionally inactivated *agfA* gene fragment is cut from pKCA1 or pKCA2 using *Eco*RI and purified. *S enteritidis* 3b is transformed with the insertionally inactivated *agfA* fragment by electroporation to allow chromosomal gene

replacement by homologous recombination. The transformants are selected on media containing chloramphenicol for those transformed with *agfA*::ΩCm or on T-medium containing Congo red for the transformants containing *agfA*::LKC1. The genotype and phenotype of the *agfA*⁻ transformants are confirmed by PCR and DNA sequencing or Western blot analysis, respectively.

This method can also be applied using cloned and sequenced chromosomal DNA fragments from *S. enteritidis* which encode larger *agfA* flanking sequences to promote higher frequency of gene replacement by homologous recombination.

EXAMPLE 13

Creation of a *sefD* (SEF18) mutant

pSCC carries an *S. enteritidis sefD* gene fragment flanked *Eco*RI and *Pst*I sites. There are no unique internal restriction sites in *sefD*, so PCR is used to generate an internal *Bam*HI site by site-specific mutagenesis. The primers KC6 and KC48 are used to amplify by PCR the 263 bp fragment. The product is digested with *Eco*RI and *Bam*HI, purified and cloned into *Eco*RI/*Bam*HI digested pUC18 to generate pKCD1. A second set of primers, KC5 and KC47, are used to amplify the 333 bp fragment. The product is digested with *Bam*HI and *Pst*I,purified, and cloned into *Bam*HI/*Pst*I digested pKCD1 to generate pKCD2. The 2.0 kb *Bam*HI ΩSm/Spc fragment from pHP45Ω (which contains streptomycin and spectinomycin resistance genes) is cloned into *Bam*HI digested pKCD2 to generate pKCD3. The *Eco*RI/*Bam*HI fragment of pKCD3 is purified prior to electroporation. The *sefD* gene containing the engineered *Bam*HI site is also insertionally inactivated using LKC2 linkers converted to LKC4 by digestion with *Bgl*II.

S enteritidis 3b is transformed with the resulting insertionally inactivated *sefD* fragment, *sefD*::ΩSmSpc, or *sefD*::LKC4 by electroporation to permit chromosomal gene replacement. The transformants are selected on media containing spectinomycin and streptomycin for those transformed with *sefD*::ΩSpcSm or screened for the inability to produce SefD for those transformed with the *sefD*::LKC4. The genotype and phenotype of the *sefD*⁻ transformants are confirmed by PCR and DNA sequencing or Western blot analysis, respectively.

This method can also be applied using cloned and sequenced chromosomal DNA fragments from *S. enteritidis* which encode larger flanking

sequences around *sefD* and therefore increase the frequency of homologous gene replacement.

EXAMPLE 14

Creation of a *FimA* (SEF21) mutant

A 171 bp *Hin*PI fragment of *fimA* previously subcloned into the unique *Acc*I site of M13mp18 is cut from the replicative form by double digestion with *Bam*HI and *Hin*dIII. This *fimA* fragment is subcloned into *Bam*HI and *Hin*dIII digested pUC18 to generate pKCF1. Two complementary oligonucleotides, LKC2A and LKC2B (Table 1), are synthesized, annealed together and then phosphorylated with T4 polynucleotide kinase to generate the double-stranded linker LKC2 which contains an internal *Bgl*II site and possessed sticky ends compatible with those generated by *Bst*BI digestion. pKCF1 is digested with *Bst*BI, purified by agarose gel electrophoresis and Sephaglas™ treatment, ligated to LKC2, digested with *Bgl*II and re-purified. pKCF1 carrying the LKC2 linker is self-ligated to generate pKCF2. The *fimA*::LKC2 fragment is cut from pKCF2 with *Bam*HI and *Hin*dIII and purified. To generate a *fimA* fragment insertionally inactivated with ΩTc, *Bst*BI, linearized fragments of pKCF1 and 1.9 kb *Hin*dIII fragments of pHP45ΩTc (carrying the tetracycline resistance gene (ΩTc)) are blunt-ended by treatment with the Klenow fragment of DNA polI and ligated to create pKCF3. Prior to introduction into *S. enteritidis*, the *fimA*::ΩTc fragment is cut from pKCF3 with *Eco*RI and *Hin*dIII and purified.

S enteritidis 3b is transformed with the resulting insertionally inactivated *fimA* fragment by electroporation to allow chromosomal *fimA* gene replacement by homologous recombination. The fimA$^-$ transformants are selected on media containing tetracycline to select cells transformed with *fimA*::ΩTc or screened for the inability to produce FimA or SEF21 for those transformed with the *fimA*::LKC2. The genotype and phenotype of the *fimA*$^-$ recombinants are confirmed by PCR and DNA sequencing or Western blot analysis, respectively.

This method can also be applied using cloned and sequenced chromosomal DNA fragments from *S. enteritidis* which encode larger flanking sequences around *fimA* and therefore increase the frequency of homologous gene replacement.

Table 1

Oligonucleotide sequences for cloning and mutagenesis procedures of *S. enteritidis* fimbrial genes.

Linker and primer names	Sequence	
LKC1[a]		
LKC1A	5'GCTAACAGAGTAAGATCTTGCTAACGAGCGG	(SEQ. ID NO.___)
LKC1B	CCGATTGTCTCATTCTAGAACGATTGCTCGC 5'	(SEQ. ID NO.___)
LKC2[a]		
LKC2A	5'CGCTAAGCAGATCTAAACCCTAATCCC	(SEQ. ID NO.___)
LKC2B	GATTCGTCTAGATTTGGGATTAGGGGC 5'	(SEQ. ID NO.___)
LKC3[a]		
LKC3A	5'GGCCTAATGACCTAAGATCTTGCTAACGAAT	(SEQ. ID NO.___)
LKC3B	GATTACTGGATTCTAGAACGATTGCTTACCG 5'	(SEQ. ID NO.___)
KC5	5'GAGAGGGAAAAAGGATCCTCATTAGTTCAAG	(SEQ. ID NO.___)
KC6	CTCTCCCTTTTTCCTAGGAGTAATCAAGTTC 5'	(SEQ. ID NO.___)
KC48	5'AGCGGATAACAATTTCACACAGGAAAC	(SEQ. ID NO.___)
KC47	5'CGCCAGGGTTTTCCCAGTCACGAC	(SEQ. ID NO.___)
LKC4[a]		
	5'GATCTAAACCCTAATCCCCGCTAAGCA	(SEQ. ID NO.___)
	ATTTGGGATTAGGGGCGATTCGTCTAG 5'	(SEQ. ID NO.___)

[a] These linkers were designed to have TAA or TAG stop codons in all three frames so that insertion of this piece of DNA into a gene will cause premature termination of the protein as well as introducing a useful cloning site.

EXAMPLE 15

In vivo expression of assembled fimbriae in *E. coli*

Immunogold labeling of *S. enteritidis* cells with polyclonal immune serum raised to purified SEF14 fimbriae revealed thin, filamentous organelles located on the cell surface (Figure 6A). *E. coli* cells hosting pKX1 or cos48 were shown by Western blot analysis to produce SefA.

To determine whether *E. coli* cells hosting pKX1 or cos48 were assembling intact fimbriae, *E. coli* cells were grown in 2.5 ml of colonization factor antigen static broth culture (Evans et al., "Hemagglutination of human group A erythrocytes by enterotoxigenic *Escherichia coli* isolated from adults with diarrhea: correlation with colonization factor," *Infect. Immun. 18*:330-337, 1977) for 60 h at 37°C. Cells were then mounted on formvar-coated grids, incubated with immune serum to native SEF 14 fimbriae, then labeled with protein A-gold (Auroprobe, Pharmacia, Uppsala, Sweden). Labeled cells were then negatively stained with ammonium molybdate.

Immunoelectron microscopic examination of *E. coli* HB101 containing cos48 revealed the presence of immunogold-labeled SEF 14 fimbriae on the cells surface (Figure 6B). However, examination of *E. coli* JM109 carrying pKX1 revealed the formation of surface blebs that were specifically gold labeled (Figure 6C). *E. coli* carrying a 44 kb fragment of *S. enteritidis* DNA (cos48) encompassing the *sefU$_2$U$_1$ABCD* operon, as well as *E. coli* carrying a 10 kb *Kpn*I fragment subcloned from cos48, displayed intact fimbriae. Conversely, *E. coli* carrying a 5.3 kb *Hin*dIII fragment subcloned from cos48 showed that distinguishable SEF14 filamentous fimbriae were not assembled.

From the foregoing, it will be appreciated that, although specific embodiments of the invention have been described herein for purposes of illustration, various modifications may be made without deviating from the spirit and scope of the invention. Accordingly, the invention is not limited except as by the appended claims.

Claims

1. A composition able to elicit an immune response in an animal comprising an isolated protein selected from the group consisting of a SefC protein, a SefD protein, a TctC protein and an AgfA protein, in combination with a physiologically acceptable carrier or diluent.

2. A vector construct comprising a mutant gene selected from the group consisting of a mutant *tctA* gene, a mutant *tctB* gene and a mutant *tctC* gene, said mutant gene capable of inactivating the corresponding tricarboxylic acid transport pathway in *Salmonella* upon incorporation of said vector construct into said *Salmonella*, to yield an attenuated *Salmonella*.

3. A biologically pure attenuated *Salmonella* comprising a mutation in a gene selected from the group consisting a *tctA* gene, a *tctB* gene, and a *tctC* gene, to yield said attenuated *Salmonella*, wherein said mutation eliminates the function of the corresponding tricarboxylic acid transport system of said attenuated *Salmonella*.

4. A composition able to elicit an immune response in an animal comprising an attenuated *Salmonella* having an inactivating mutation in a gene or operon selected from the group consisting of a *tctA* gene, a *tctB* gene, a *tctC* gene, a *tctI* operon, a *tctII* operon and a *tctIII* operon, in combination with a physiologically acceptable carrier or diluent.

5. A composition able to elicit an immune response in an animal comprising an attenuated *Salmonella* having a mutation in two or more fimbriae encoding genes selected from the group consisting of *sefA*, *sefD*, *agfA*, and *fimA*, wherein said mutations effectively prevent production of fimbriae from said genes, in combination with a physiologically acceptable carrier or diluent.

6. The composition of any one of claim 4 or 5 able to elicit an immune response in an animal wherein said attenuated *Salmonella* is capable of expressing a foreign antigen in one r more of its fimbriae or in an aggregate.

7. The *Salmonella* of claim 6 wherein said attenuated *Salmonella* is capable of expressing a foreign antigen fused to a SefA, SefC, SefD, TctC, FimA, or AgfA protein.

8. An expression vector construct comprising a *sefA* gene, a *sefC* gene, *sefD* gene, a *FimA* gene, a *tctC* gene or an *agfA* gene operably linked in open reading frame to a foreign gene to yield a dicistronic gene product, said dicistronic gene product capable of being expressed in a fimbriae, or an aggregate comprising said gene product, of a *Salmonella*.

9. An expression vector construct comprising a *sefA* gene, a *sefC* gene, *sefD* gene, a *fimA* gene, a *tctC* gene or an *agfA* gene operably linked in open reading frame to a foreign gene to yield a dicistronic gene product, said dicistronic gene product capable of being expressed in a fimbriae, or an aggregate comprising said gene product, of an *E. coli*.

10. An expression vector construct comprising an *agfA* gene capable of producing in *E. coli* a stable fimbriae comprising AgfA protein.

11. A stable fimbriae comprising an AgfA protein fused to one or more foreign antigens.

12. A method of eliciting an immune response in an animal, comprising:

(a) separating fimbriae comprising a SefA protein, a SefD protein, a FimA protein, or an AgfA protein fused to a foreign antigen grown on a *Salmonella* host cell from said *Salmonella* host cell; and

(b) introducing said fimbriae into said animal in conjunction with a physiologically acceptable carrier or diluent.

13. A method of eliciting an immune response in an animal to *Salmonella*, comprising:

(a) separating an amino acid polymer comprising a SefA protein, a SefC protein, a SefD protein, a FimA protein, a TctC protein, or an AgfA protein fused to a foreign antigen grown on a *Salmonella* host cell from said *Salmonella* host cell; and

(b) introducing said amino acid polymer into said animal in conjunction with a physiologically acceptable carrier or diluent.

14. A method of eliciting an immune response in an animal to *Salmonella*, comprising:

(a) separating fimbriae comprising a SefA protein, a SefD protein or an AgfA protein grown on an *E. coli* host cell from said *E. coli* host cell; and

(b) introducing said fimbriae into said animal in conjunction with a physiologically acceptable carrier or diluent.

15. A method of eliciting an immune response in an animal to *Salmonella*, comprising:

(a) separating an amino acid polymer comprising a SefA protein, a SefC protein, a SefD protein, a TctC protein, or an AgfA protein grown on an *E. coli* host cell from said *E. coli* host cell; and

(b) introducing said amino acid polymer into said animal in conjunction with a physiologically acceptable carrier or diluent.

16. The method of claim 14 or 15 wherein said protein is fused to a foreign antigen.

17. A method of eliciting an immune response in an animal to *Salmonella* comprising introducing into said animal one of the group consisting of an isolated SefD protein, an isolated SefC protein and an isolated AgfA protein, in combination with a physiologically acceptable carrier or diluent.

18. A method of eliciting an immune response in an animal to *Salmonella*, comprising introducing an attenuated *Salmonella* into said animal, said *Salmonella* comprising an ineffective gene or operon selected from the group consisting of a *tctA* gene, a *tctB* gene, a *tctC* gene, a *tctI* operon, a *tctII* operon, and a *tctIII* operon.

19. A method of eliciting an immune response in an animal comprising introducing into an animal an attenuated *Salmonella* having a mutation in two or more fimbriae encoding genes selected from the group consisting of a *sefA*, *sefD*, *agfA*, and *fimA*, wherein said mutations effectively prevent production of fimbriae from said genes, in combination with a physiologically acceptable carrier or diluent.

20. The method of claim 18 or 19 wherein said attenuated *Salmonella* expresses one or more foreign antigens, said foreign antigens eliciting an immune response from said animal.

21. The method of claim 20 wherein said one or more foreign antigens are located on a fimbriae of said attenuated *Salmonella*.

22. A method of eliciting an immune response in an animal to *Salmonella* comprising introducing an *E. coli* into said animal, said *E. coli* expressing *Salmonella* fimbriae.

23. The method of claim 22 wherein said *Salmonella* fimbriae comprise *agfA* gene product.

24. The method of claim 22 wherein said *Salmonella* fimbriae further comprises a foreign antigen, said foreign antigen capable of eliciting an immune response from said animal.

25. A method of eliciting an immune response in an animal to *Salmonella* comprising introducing a vector into said animal wherein said vector comprises a gene selected from the group consisting of a *sefC* gene, *sefD* gene, *tctC*, and an *agfA* gene.

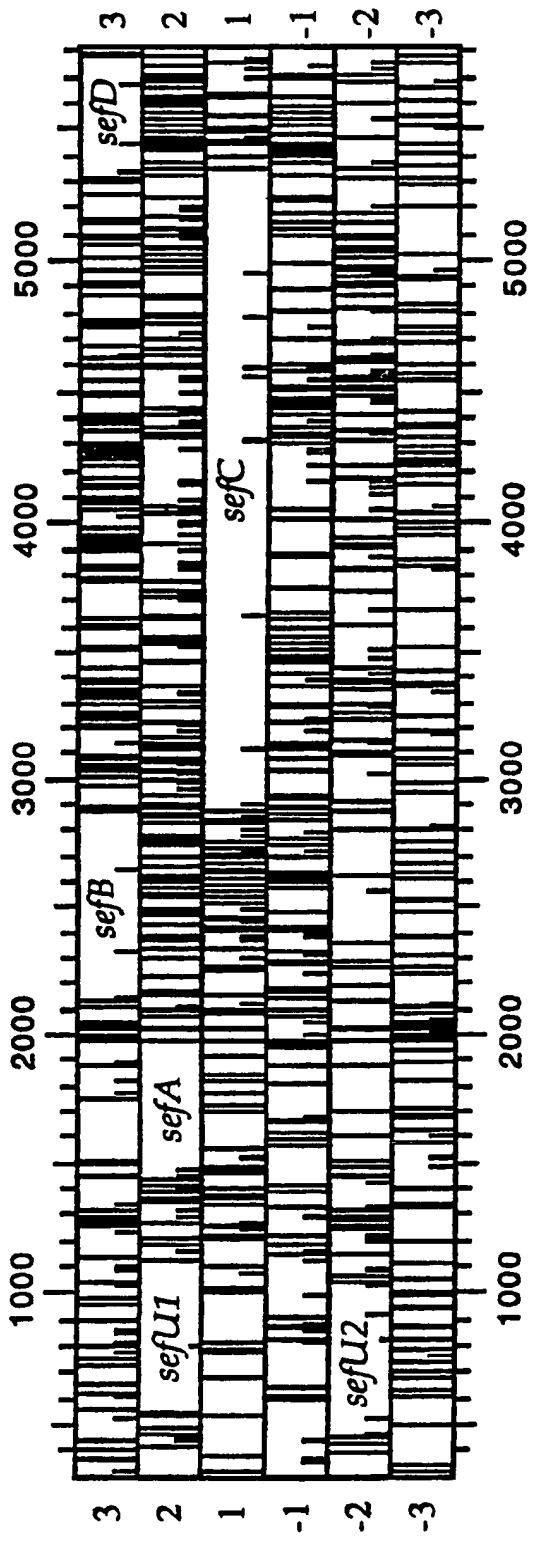
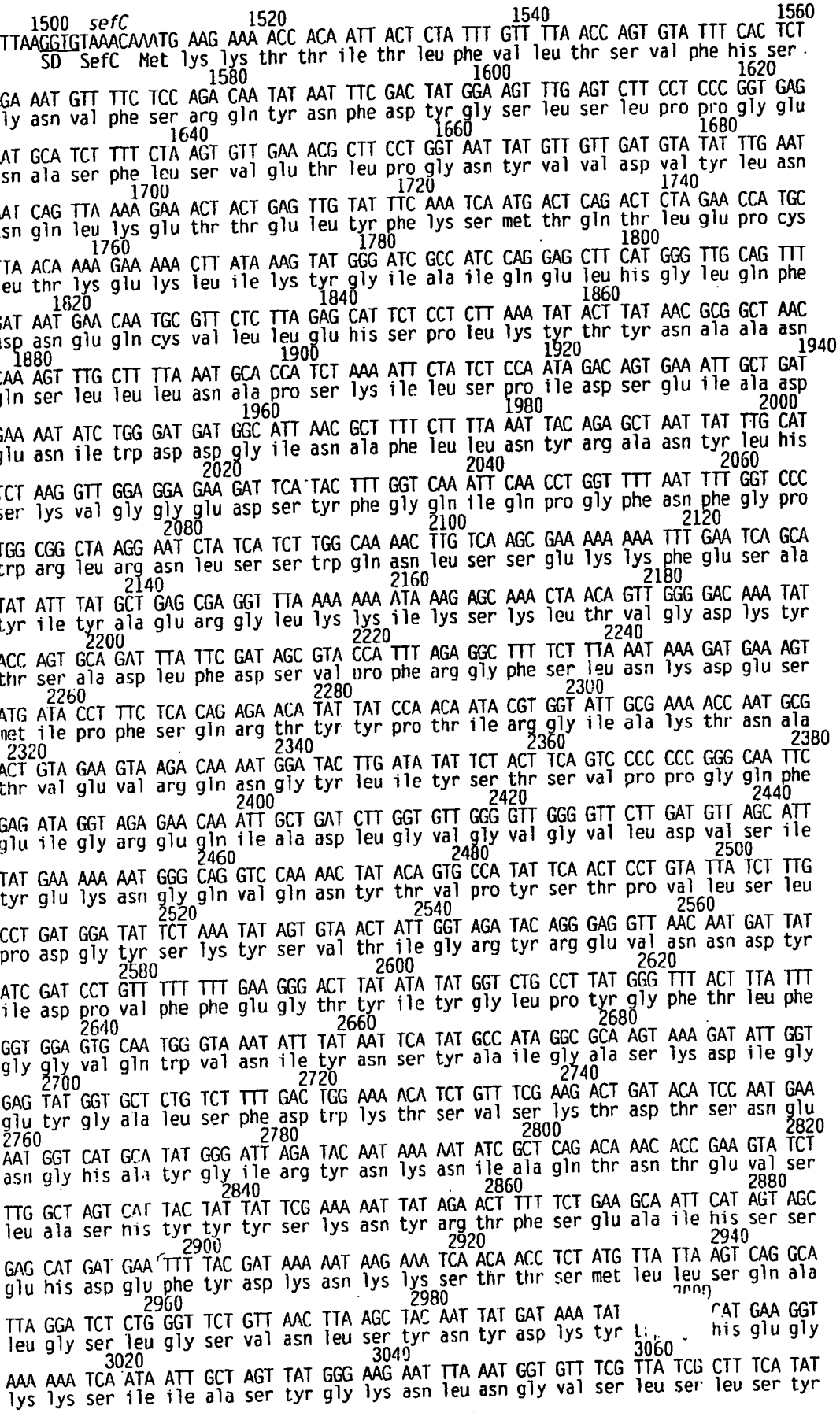

3
2
1
-1
-2
-3
sefD
sefC
sefB
sefA
sefU1
sefU2
5000
4000
3000
2000
1000

FIG. 1

FIGURE 2A

```
                    20                  40                  60                  80
GGGGATGTTGTGTAAAGATAAAAAAATAGTGATCCTTGTTTTTTTTCTTAAATTTTTAAAATGGCGTGAGTATATTAGCATCC
                  100                 120      sefA       140          -10        *
GCACAGATAAATTGTGCGAATGCTAATAGTTGATTTTTGGAGATTTTGTAAT ATG CGT AAA TCA GCA TCT GCA GTA
*       *      **       IHF                 SD    SefA    Met arg lys ser ala ser ala val
160                 180                 200                 220
GCA GTT CTT GCT TTA ATT GCA TGT GGC AGT GCC CAC GCA GCT GGC TTT GTT GGT AAC AAA GCA
ala val leu ala leu ile ala cys gly ser ala his ala ala gly phe val gly asn lys ala
                240                 260                 280
GTG GTT CAG GCA GCG GTT ACT ATT GCA GCT CAG AAT ACA ACA TCA GCC AAC TGG AGT CAG GAT
val val gln ala ala val thr ile ala ala gln asn thr thr ser ala asn trp ser gln asp
              300                 320                 340
CCT GGC TTT ACA GGG CCT GCT GTT GCT GCT GGT CAG AAA GTT GGT ACT CTC AGC ATT ACT GCT
pro gly phe thr gly pro ala val ala ala gly gln lys val gly thr leu ser ile thr ala
            360                 380                 400
ACT GGT CCA CAT AAC TCA GTA TCT ATT GCA GGT AAA GGG GCT TCG GTA TCT GGT GGT GTA GCC
thr gly pro his asn ser val ser ile ala gly lys gly ala ser val ser gly gly val ala
          420                 440                 460
ACT GTC CCG TTC GTT GAT GGA CAA GGA CAG CCT GTT TTC CGT GGG CGT ATT CAG GGA GCC AAT
thr val pro phe val asp gly gln gly gln pro val phe arg gly arg ile gln gly ala asn
        480                 500                 520
ATT AAT GAC CAA GCA AAT ACT GGA ATT GAC GGG CTT GCA GGT TGG CGA GTT GCC AGC TCT CAA
ile asn asp gln ala asn thr gly ile asp gly leu ala gly trp arg val ala ser ser gln
      540                 560                 580                 600
GAA ACG CTA AAT GTC CCT GTC ACA ACC TTT GGT AAA TCG ACC CTG CCA GCA GGT ACT TTC ACT
glu thr leu asn val pro val thr thr phe gly lys ser thr leu pro ala gly thr phe thr
                620
GCG ACC TTC TAC GTT CAG CAG TAT CAA AAC TAA
ala thr phe tyr val gln gln tyr gln asn XXX

      640                 660                 680                 700
TTTAATTTAAACTTTATAAATGCCCTCAATATGAGGGCATTTGGATAATTTTATTATTTTAAAAATATCTATTTTGAATAGATA
    720                 740      sefB       760                 780
GGTTTTATGCTTCCATGCAAAAACTTAAAGAGGGATT   ATG TAT ATT TTG AAT AAA TTT ATA CGT AGA ACT
                              SD   SefB Met tyr ile leu asn lys phe ile arg arg thr
              800                 820                 840
GTT ATC TTT TTC TTT TTT TGC TAC CTT CCA ATT GCT TCT TCG GAA AGT AAA AAA ATT GAG CAA
val ile phe phe phe phe cys tyr leu pro ile ala ser ser glu ser lys lys ile glu gln
            860                 880                 900
CCA TTA TTA ACA CAA AAA TAT TAT GGC CTA AGA TTG GGC ACT ACA CGT GTT ATT TAT AAA GAA
pro leu leu thr gln lys tyr tyr gly leu arg leu gly thr thr arg val ile tyr lys glu
          920                 940                 960
GAT GCT CCA TCA ACA AGT TTT TGG ATT ATG AAT GAA AAA GAA TAT CCA ATC CTT GTT CAA ACT
asp ala pro ser thr ser phe trp ile met asn glu lys glu tyr pro ile leu val gln thr
        980                1000                1020
CAA GTA TAT AAT GAT GAT AAA TCA TCA AAA GCT CCA TTT ATT GTA ACA CCA CCT ATT TTG AAA
gln val tyr asn asp asp lys ser ser lys ala pro phe ile val thr pro pro ile leu lys
      1040                1060                1080                1100
GTT GAA AGT AAT GCG CGA ACA AGA TTG AAG GTA ATA CCA ACA AGT AAT CTA TTC AAT AAA AAT
val glu ser asn ala arg thr arg leu lys val ile pro thr ser asn leu phe asn lys asn
              1120                1140                1160
GAG GAG TCT TTG TAT TGG TTG TGT GTA AAA GGA GTC CCA CCA CTA AAT GAT AAT GAA AGC AAT
glu glu ser leu tyr trp leu cys val lys gly val pro pro leu asn asp asn glu ser asn
            1180                1200                1220
AAT AAA AAC AAC ATA ACT ACG AAT CTT AAT GTG AAT GTG GTT ACG AAT AGT TGT ATT AAA TTA
asn lys asn asn ile thr thr asn leu asn val asn val val thr asn ser cys ile lys leu
          1240                1260                1280
ATT TAT AGG CCT AAA ACT ATA GAC TTA ACG ACA ATG GAG ATT GCA GAT AAA TTA AAG TTA GAG
ile tyr arg pro lys thr ile asp leu thr thr met glu ile ala asp lys leu lys leu glu
        1300                1320                1340
AGA AAA GGA AAT AGT ATA GTT ATA AAG AAT CCA ACA TCA TCA TAT GTG AAT ATT GCA AAT ATT
arg lys gly asn ser ile val ile lys asn pro thr ser ser tyr val asn ile ala asn ile
      1360                1380                1400
AAA TCT GGT AAT TTA AGT TTT AAT ATT CCA AAT GGA TAT ATT GAG CCA TTT GGA TAT GCT CAA
lys ser gly asn leu ser phe asn ile pro asn gly tyr ile glu pro phe gly tyr ala gln
    1420                1440                1460                1480
TTA CCT GGT GGA GTA CAT AGT AAA ATA ACT TTG ACT ATT TTG GAT GAT AAC GGC GCT GAA ATT
leu pro gly gly val his ser lys ile thr leu thr ile leu asp asp asn gly ala glu ile

ATA AGA GAT TAT TAG
ile arg asp tyr XXX
```

FIGURE 2B



ISA/EP

FIGURE 2C

```
      3080                        3100                        3120
ACG AAA AGT ACA TCA AAG ATT AGT GAA GAA AAT GAA GAT TTA TTC AGT TTT CTA CTC AGT GTA
thr lys ser thr ser lys ile ser glu glu asn glu asp leu phe ser phe leu leu ser val
  3140                        3160                        3180                        3200
CCT TTG CAA AAA CTT ACA AAT CAT GAA ATG TAT GCT ACA TAT CAA AAC TCA TCC TCT TCA AAG
pro leu gln lys leu thr asn his glu met tyr ala thr tyr gln asn ser ser ser ser lys
                      3220                        3240                        3260
CAT GAT ATG AAT CAT GAT TTA GGT ATT ACT GGT GTG GCA TTT AAT AGC CAA TTG ACA TGG CAA
his asp met asn his asp leu gly ile thr gly val ala phe asn ser gln leu thr trp gln
                    3280                        3300                        3320
GCA AGA GGG CAA ATA GAA GAT AAA TCG AAA AAT CAA AAG GCT ACA TTT TTA AAT GCT TCT TGG
ala arg gly gln ile glu asp lys ser lys asn gln lys ala thr phe leu asn ala ser trp
                      3340                        3360                        3380
CGA GGT ACT TAT GGG GAG ATC GGA GCA AAC TAT AGT CAT AAT GAA ATA AAT CGT GAT ATT GGG
arg gly thr tyr gly glu ile gly ala asn tyr ser his asn glu ile asn arg asp ile gly
                  3400                        3420                        3440
ATG AAT GTT TCT GGT GGG GTG ATT GCT CAT TCA TCA GGA ATT ACG TTT GGT CAG AGT ATA TCG
met asn val ser gly gly val ile ala his ser ser gly ile thr phe gly gln ser ile ser
                3460                        3480                        3500
GAT ACT GCT GCA CTG GTA GAG GCT AAA GGT GTA AGT GGG GCA AAA GTT CTG GGC CTA CCA GGT
asp thr ala ala leu val glu ala lys gly val ser gly ala lys val leu gly leu pro gly
      3520                        3540                        3560
GTT AGA ACC GAT TTT AGA GGC TAT ACA ATA TCC AGT TAT CTT ACT CCA TAT ATG AAT AAC TTC
val arg thr asp phe arg gly tyr thr ile ser ser tyr leu thr pro tyr met asn asn phe
 3580                        3600                        3620                        3640
ATA TCT ATA GAT CCA ACA ACG TTA CCA ATA AAT ACG GAT ATT AGG CAA ACT GAT ATT CAA GTA
ile ser ile asp pro thr thr leu pro ile asn thr asp ile arg gln thr asp ile gln val
                      3660                        3680                        3700
GTT CCT ACC GAA GGT GCT ATT GTA AAA GCT GTA TAT AAA ACA AGC GTG GGT ACT AAT GCA TTA
val pro thr glu gly ala ile val lys ala val tyr lys thr ser val gly thr asn ala leu
                    3720                        3740                        3760
ATT AGA ATT ACA AGA ACA AAT GGA AAG CCA CTA GCT CTT GGC ACA GTT CTT TCA CTT AAG AAT
ile arg ile thr arg thr asn gly lys pro leu ala leu gly thr val leu ser leu lys asn
                      3780                        3800                        3820
AAT GAT GGA GTA ATC CAA TCA ACA TCT ATT GTT GGC GAA GAT GGT CAG GCA TAT GTA TCT GGA
asn asp gly val ile gln ser thr ser ile val gly glu asp gly gln ala tyr val ser gly
                  3840                        3860                        3880
TTG TCA GGA GTG CAA AAA TTA ATC GCT TCG TGG GGG AAT AAG CCC TCC GAT ACT TGT ACA GTT
leu ser gly val gln lys leu ile ala ser trp gly asn lys pro ser asp thr cys thr val
      3900                        3920                        3940
TTT TAC TCT CTT CCC GAT AAA AAC AAA GGT CAG ATT AGC TTT TTA AAT GGA GTG TGC AAA TGA
phe tyr ser leu pro asp lys asn lys gly gln ile ser phe leu asn gly val cys lys XXX
```

FIGURE 2D

sefD

```
          3960                                3980
ATG AAT CAG TAT AAT TCG TCA ATA CCT AAG TTC ATT GTC TCT GTT
met asn gln tyr asn ser ser ile pro lys phe ile val ser val
```

sefD

```
  4000                                4020                                4040
TTT CTG ATT GTT ACT GGT TTT TTC AGC TCA ACT ATT AAA GCA CAA
phe leu ile val thr gly phe phe ser ser thr ile lys ala gln

                          4060                                4080
GAA CTT AAA TTA ATG ATT AAA ATA AAT GAG GCT GTT TTT TAT GAC
glu leu lys leu met ile lys ile asn glu ala val phe tyr asp

                  4100                                4120
CGT ATT ACA AGT AAT AAA ATA ATA GGT ACG GGG CAT CTA TTT AAC
arg ile thr ser asn lys ile ile gly thr gly his leu phe asn

          4140                                4160
AGA GAG GGA AAA AAA ATC CTC ATT AGT TCA AGT TTA GAA AAA ATT
arg glu gly lys lys ile leu ile ser ser ser leu glu lys ile

  4180                                4200                                4220
AAA AAT ACC CCA GGG GCA TAT ATT ATT AGA GGT CAG AAT AAC TCA
lys asn thr pro gly ala tyr ile ile arg gly gln asn asn ser

                          4240                                4200
GCC CAT AAG CTT AGG ATA AGA ATA GGT GGA GAA GAC TGG CAA CCA
ala his lys leu arg ile arg ile gly gly glu asp trp gln pro

                  4280                                4300
GAT AAT TCA GGT ATT GGT ATG GTA TCT CAT TCT GAT TTT ACT AAT
asp asn ser gly ile gly met val ser his ser asp phe thr asn

          4320                                4340
GAA TTT AAT ATT TAT TTT TTT GGG AAT GGA GAC ATT CCT GTT GAC
glu phe asn ile tyr phe phe gly asn gly asp ile pro val asp

  4360                                4380                                4400
ACA TAT TTA ATA AGC ATA TAT GCG ACA GAA ATT GAA TTA TAA TAA
thr tyr leu ile ser ile tyr ala thr glu ile glu leu XXX XXX
```

FIGURE 3A

478/11

TC ACC CAC CCA TTT CTG ATT CGG GCC ACT GGC GTA AAA GCC CTG CTT CAG CAG ATT CTC
AG TGG GTG GGT AAA GAC TAA GCC CGG TGA CCG CAT TTT CGG GAC GAA GTC GTC TAA GAG

OPA gly gly met glu ser glu pro trp gln arg leu leu gly ala glu ala ser glu arg

508/21 538/31

val lys leu ser pro

sefU1—>

TGG ACT GGC AGA CCA TGT TCG CGG TAA CTG ACT GGA CTG ATC TTC CGT GAA GCT TTC GCC
ACC TGA CCG TCT GGT ACA AGC GCC ATT GAC TGA CCT GAC TAG AAG GCA CTT CGA AAG CGG

ser gln cys val met asn ala thr val ser gln val ser arg gly his leu lys arg gly

568/41 598/51

ala ala leu pro ala gln ala gly arg tyr gly phe tyr val ile his pro ser leu ser

CGC AGC ACT GCC GGC GCA GGC TGG CCG CTA CGG TTT TTA TGT TAT ACA CCC GTC CCT GAG
GCG TCG TGA CGG CCG CGT CCG ACC GGC GAT GCC AAA AAT ACA ATA TGT GGG CAG GGA CTC

cys cys gln arg arg leu ser ala ala val thr lys ile asn tyr val arg gly gln ala

628/61 658/71

thr lys leu ile arg gln ala trp arg thr val ala leu phe cys val thr glu cys leu

CAC GAA GCT CAT CCG TCA GGC GTG GCG TAC CGT AGC GCT GTT TTG CGT CAC TGA ATG CCT
GTG CTT CGA GTA GGC AGT CCG CAC CGC ATG GCA TCG CGA CAA AAC GCA GTG ACT TAC GGA

arg leu glu asp thr leu arg pro thr gly tyr arg gln lys ala asp ser phe ala glu

688/81 718/91

pro tyr asp val ile thr asp lys ser glu leu leu thr pro asp val pro ala val thr

CCC GTA CGA CGT TAT CAC AGA CAA GTC GGA ACT GCT GAC GCC GGA CGT ACC AGC TGT TAC
GGG CAT GCT GCA ATA GTG TCT GTT CAG CCT TGA CGA CTG CGG CTT GCA TGG TCG ACA ATG

arg val val asn asp cys val leu arg phe gln gln arg arg val tyr trp ser asn arg

748/101 778/111

gly asn leu lys tyr thr ala tyr gly phe asp thr glu leu ser leu met phe phe asp

GGG CAA CCT GAA GTA CAC GGC ATA TGG CTT TGA TAC TGA ACT CAG CCT GAT GTT TTT CGA
CCC GTT GGA CTT CAT GTG CCG TAT ACC GAA ACT ATG ACT TGA GTC GGA CTA CAA AAA GCT

ala val gln leu val arg cys ile ala lys ile ser phe glu ala gln his lys glu ile

808/121 838/131

glu asp ile leu his phe arg arg phe ala lys tyr val ala thr ile leu glu asn gly

TGA AGA CAT ACT TCA TTT CAG GCG TTT CGC GAA GTA TGT CGC GAC CAT TCT GGA GAA TGG
ACT TCT GTA TGA AGT AAA GTC CGC AAA GCG CTT CAT ACA GCG CTG GTA AGA CCT CTT ACC

phe val tyr lys met glu pro thr glu arg leu ile asp arg gly asn gln leu ile thr

FIGURE 3B

868/141 898/151
gln phe leu ile pro phe cys gln leu thr leu gln thr asp asp phe cys gly his leu

TCA GTT CCT CAT CCC GTT CTG CCA GTT GAC GCT TCA GAC GGA CGA TTT CTG CGG ACA TCT
AGT CAA GGA GTA GGG CAA GAC GGT CAA CTG CGA AGT CTG CCT GCT AAA GAC GCC TGT AGA

leu glu glu asp arg glu ala leu gln arg lys leu arg val ile glu ala ser met glu

928/161 958/171
leu phe ala phe arg arg arg glu leu ile leu leu phe ala ser pro val val glu leu

CCT GTT CGC GTT CAG AAG AAG AGA GCT GAT TTT GCT GTT TGC TTC GCC AGT TGT AGA GCT
GGA CAA GCG CAA GTC TTC TTC TCT CGA CTA AAA CGA CAA ACG AAG CGG TCA ACA TCT CGA

gln glu arg glu ser ser ser leu gln asn gln gln lys ser arg trp asn tyr leu gln

988/181 1018/191
arg phe ile gln val lys leu ala gly ser arg ser his pro asn ala leu ala ser phe

GCG ATT CAT ACA GGT TAA GCT CGC TGG TAG CCG CAG CCA TCC CAA TGC GTT AGC CAG TTT
CGC TAA GTA TGT CCA ATT CGA GCG ACC ATC GGC GTC GGT AGG GTT ACG CAA TCG GTC AAA
← *sefU2*

ser glu tyr leu asn leu glu ser thr ala ala ala met

1048/201 1078/211
lys ala ser ser arg asn ser gly val cys cys leu his gly phe leu val val asp ala

CAA GGC TTC GTC GCG AAA TTC AGG CGT ATG TTG CTT GCA TGG CTT TTT GGT GGT TGA TGC
GTT CCG AAG CAG CGC TTT AAG TCC GCA TAC AAC GAA CGT ACC GAA AAA CCA CCA ACT ACG

1108/221
ala phe val met OPA

TGC TTT TGT CAT GTG A
ACG AAA ACA GTA CAC T

FIGURE 4A

```
      3303        3312        3321        3330        3339        3348
ACC GGG GTT TAT CGC TTT ACC TTT GAC AGC GTT CAT CTT TCC GAC GGC GTA CAG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Arg Gly Leu Ser Leu Tyr Leu *** Gln Arg Ser Ser Phe Arg Arg Arg Thr Val

      3357        3366        3375        3384        3393        3402
TTT ATC GTC GTG GTG ATC GGC CTG TTC TCG GTA TCA GAA ATA CTT TTA ATG CTG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Tyr Arg Arg Gly Asp Arg Pro Val Leu Gly Ile Arg Asn Thr Phe Asn Ala Gly

      3411        3420        3429        3438        3447        3456
GAA CAT ACC AGC AGC GGG CAA ACA ATG GTC CGC AAA ACG GGT CGA ATG TTG TTC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Thr Tyr Gln Gln Arg Ala Asn Asn Gly Pro Gln Asn Gly Ser Asn Val Val Gln

      3465        3474        3483        3492        3501        3510
AAC CTG AAA GAA GGC GCG CAG TGT ATC GGC ACC ACC CTG CGT TCT TCG GTA ATC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Pro Glu Arg Arg Arg Ala Val Tyr Arg His His Pro Ala Phe Phe Gly Asn Arg

      3519        3528        3537        3546        3555        3564
GGC TTT TTT GTC GGC GTA TTG CCC GGC GCC GGG CGA CCA TTG CCA GCG CCA TTA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Leu Phe Cys Arg Arg Ile Ala Arg Arg Arg Ala Thr Ile Ala Ser Ala Ile Thr

      3573        3582        3591        3600        3609        3618
CCT ATA TGA CCG AGA AAA AAC TCA GCG GCA ACA GCG ATA GCT TCG GCA AAG GGG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Tyr Met Thr Glu Lys Lys Leu Ser Gly Asn Ser Asp Ser Phe Gly Lys Gly Asp

      3627        3636        3645        3654        3663        3672
ATA TTC GCG GCG TCG CGG CGC CGG AGG CGG CAA ACA ACG CCT CTG CCT GCG GCT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Ile Arg Gly Val Ala Ala Pro Glu Ala Ala Asn Asn Ala Ser Ala Cys Gly Ser

      3681        3690        3699        3708        3717        3726
CCT TCA TCC CGA TGC TGA CGC TGG GCG TTC CCG GTT CCG GCA CTA CGG CAG TGA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Phe Ile Pro Met Leu Thr Leu Gly Val Pro Gly Ser Gly Thr Thr Ala Val Met

      3735        3744        3753        3762        3771        3780
TGA TGG GGG CGC TGA CGC TGT ACA ACA TCA CGC CAG GCC CGG CGA TGT TCA CCG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Met Gly Ala Leu Thr Leu Tyr Asn Ile Thr Pro Gly Pro Ala Met Phe Thr Glu

      3789        3798        3807        3816        3825        3834
AAC AGC CGG ATA TCG TCT GGG GAC TCA TCG CTG CGC TGC TGA TTG CGA ACG TGA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Gln Pro Asp Ile Val Trp Gly Leu Ile Ala Ala Leu Leu Ile Ala Asn Val Met

      3843        3852        3861        3870        3879        3888
TGC TGC TGA TCA TGA ATA TCC CGT TGA TCG GTC TGT TCA CCC GTA TGC TCA CCA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Leu Leu Ile Met Asn Ile Pro Leu Ile Gly Leu Phe Thr Arg Met Leu Thr Ile
```

FIGURE 4B

```
      3897        3906        3915        3924        3933        3942
TTC CGC TGT GGT TCC TGG TAC CCG CCA TCG CTG CCG TAT CGG CGG TGG GGG TGT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Pro Leu Trp Phe Leu Val Pro Ala Ile Ala Ala Val Ser Ala Val Gly Val Tyr

      3951        3960        3969        3978        3987        3996
ATG CGG TAC ACA GCA CCA CCT TCG ATC TGG TGC TGA TGG TCG CGC TCG GCG TGT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Ala Val His Ser Thr Thr Phe Asp Leu Val Leu Met Val Ala Leu Gly Val Leu

      4005        4014        4023        4032        4041        4050
TAG GGT ACA TTT TAC GTA AAA TGC ACT TCC CCA TGT CAC CGC TGA TCC TGG GGT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Gly Tyr Ile Leu Arg Lys Met His Phe Pro Met Ser Pro Leu Ile Leu Gly Phe

      4059        4068        4077        4086        4095        4104
TCG TAC TGG GGG AAA TGC TGG AGC AGA ACC TGC GTC GCG CAC TCT CCA TCA GTA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Val Leu Gly Glu Met Leu Glu Gln Asn Leu Arg Arg Ala Leu Ser Ile Ser Asn

      4113        4122        4131        4140        4149        4158
ACG GCA ATA TGG CGA TTT TGT GGC AAA GCG GCG TTG CCA AAG CCC TGC TGA TCA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Gly Asn Met Ala Ile Leu Trp Gln Ser Gly Val Ala Lys Ala Leu Leu Ile Met

      4167        4176        4185        4194        4203        4212
TGG CGA TCA TGG TCA TTG TCG TAC CGC CAG TGT TAC GTC TGC TCC GTA AAC ACA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Ala Ile Met Val Ile Val Val Pro Pro Val Leu Arg Leu Leu Arg Lys His Ser

      4221        4230        4239        4248        4257        4266
GCC GTA AAC CGC AGG TTG ACG CCG GTT AAT TGA CTG CTG AAA TAC GTT GTA CTT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Arg Lys Pro Gln Val Asp Ala Gly *** Leu Thr Ala Glu Ile Arg Cys Thr Cys

      4275        4284        4293        4302        4311        4320
GTC CGG CCT ACG CGC TCA TGT GTC AGG CCG GGC ACA TCC CCG CCA GCA TTC ACT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Pro Ala Tyr Ala Leu Met Cys Gln Ala Gly His Ile Pro Ala Ser Ile His Phe

      4329        4338        4347        4356        4365        4374
TTC CCC ATA ACG CCT CTC ATT TTA CAC CCC TTC TTG CCG TTG TCA GGC TCG TGG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Pro His Asn Ala Ser His Phe Thr Pro Leu Leu Ala Val Val Arg Leu Val Ala

      4383        4392        4401        4410        4419
CGC CGT TAA CCT CAC CCT TTG CAT TGT TAA ATA TTT GTT GTT TTT G
Pro Leu Thr Ser Pro Phe Ala Leu Leu Asn Ile Cys Cys Phe ***
```

FIGURE 5

```
          2735       2744       2753       2762       2771       2780
5' ATG GAT ACC TGG ATA TAT CTT TCT CAG GGC TTT GCG GTG GCG ATG ACG CCG GAA
   Met Asp Thr Trp Ile Tyr Leu Ser Gln Gly Phe Ala Val Ala Met Thr Pro Glu

          2789       2798       2807       2816       2825       2834
   AAC CTG GTT ATC GCG TTG ATT GGC TGC TTC GTG GGC ACG ATC GTC GGT CTG CTG
   Asn Leu Val Ile Ala Leu Ile Gly Cys Phe Val Gly Thr Ile Val Gly Leu Leu

          2843       2852       2861       2870       2879       2888
   CCG GGT CTG GGA CCG ATC AAC GGC GTG GCA ATC TTA CTG CCG CTG GCC TTT GCG
   Pro Gly Leu Gly Pro Ile Asn Gly Val Ala Ile Leu Leu Pro Leu Ala Phe Ala

          2897       2906       2915       2924       2933       2942
   TTG CAT CTG CCT GCG GAG TCG GCG CTA ATT CTG CTG GCG ACG GTG TAC ATT GGC
   Leu His Leu Pro Ala Glu Ser Ala Leu Ile Leu Leu Ala Thr Val Tyr Ile Gly

          2951       2960       2969       2978       2987       2996
   TGT GAG TAT GGC GGC AGG ATC TCC TCC ATA TTG CTC AAC GTC CCC GGC GAT GCG
   Cys Glu Tyr Gly Gly Arg Ile Ser Ser Ile Leu Leu Asn Val Pro Gly Asp Ala

          3005       3014       3023       3032       3041       3050
   GCG GCG ATC ATG ACG GCG CTG GAC GGT TAC CCG ATG GCG CAG CAA GGG AAA GGC
   Ala Ala Ile Met Thr Ala Leu Asp Gly Tyr Pro Met Ala Gln Gln Gly Lys Gly

          3059       3068       3077       3086       3095       3104
   GGC GTA GCG CTG TCG ATT TCC GCA GTC AGC TCA TTT TTC GGT TCA TTA ATC GCT
   Gly Val Ala Leu Ser Ile Ser Ala Val Ser Ser Phe Phe Gly Ser Leu Ile Ala

          3113       3122       3131       3140       3149       3158
   ATC GGC GGC ATC ATT CTG TTC GCC CCT TTA CTG GCG CAA TGG TCG CTG GCC TTT
   Ile Gly Gly Ile Ile Leu Phe Ala Pro Leu Leu Ala Gln Trp Ser Leu Ala Phe

          3167       3176       3185       3194       3203       3212
   GGG CCG GCG GAA TAT TTC GCC TTA ATG GTT TTT GCC ATC GCC TGT CTT GGC AGC
   Gly Pro Ala Glu Tyr Phe Ala Leu Met Val Phe Ala Ile Ala Cys Leu Gly Ser

          3221       3230
   ATG ATG GCG CAA AAC CCG GCT TAA 3'
   Met Met Ala Gln Asn Pro Ala ***
```

FIGURE 6A

```
          1301       1310       1319       1328       1337       1346
5' ATG AAA AAA CAA TTA CTT CGT ACC CTT ACT GCA AGC ATT TTA TTA ATG AGT ACC
   Met Lys Lys Gln Leu Leu Arg Thr Leu Thr Ala Ser Ile Leu Leu Met Ser Thr

          1355       1364       1373       1382       1391       1400
   TCT GTT CTG GCG CAG GAG GCG CCG TCG CGA ACG GAA TGT ATC GCG CCA GCC AAA
   Ser Val Leu Ala Gln Glu Ala Pro Ser Arg Thr Glu Cys Ile Ala Pro Ala Lys

          1409       1418       1427       1436       1445       1454
   CCT GGC GGC GGT TTC GAC CTC ACC TGT AAG CTG ATT CAG GTG AGT TTG CTG GAG
   Pro Gly Gly Gly Phe Asp Leu Thr Cys Lys Leu Ile Gln Val Ser Leu Leu Glu

          1463       1472       1481       1490       1499       1508
   ACT GGC GCT ATC GAG AAA CCC ATG CGG GTA ACG TAT ATG CCC GGC GGC GTC GGC
   Thr Gly Ala Ile Glu Lys Pro Met Arg Val Thr Tyr Met Pro Gly Gly Val Gly

          1517       1526       1535       1544       1553       1562
   GCT GTG GCC TAT AAC GCG ATA GTC GCC CAA CGC CCT GGC GAA CCC GGG ACT GTG
   Ala Val Ala Tyr Asn Ala Ile Val Ala Gln Arg Pro Gly Glu Pro Gly Thr Val

          1571       1580       1589       1598       1607       1616
   GTC GCC TTT TCC GGC GGT TCG CTG CTC AAC CTG TCG CAG GGG AAA TTT GGT CGC
   Val Ala Phe Ser Gly Gly Ser Leu Leu Asn Leu Ser Gln Gly Lys Phe Gly Arg

          1625       1634       1643       1652       1661       1670
   TAC GGC GTG GAT GAT GTG CGC TGG CTG GCA AGC GTG GGC ACT GAT TAC GGC ATG
   Tyr Gly Val Asp Asp Val Arg Trp Leu Ala Ser Val Gly Thr Asp Tyr Gly Met

          1679       1688       1697       1706       1715       1724
   ATC GCC GTG CGT GCG GAT TCT CCG TGG AAA ACG CTG AAA GAT CTG ATG ACG GCG
   Ile Ala Val Arg Ala Asp Ser Pro Trp Lys Thr Leu Lys Asp Leu Met Thr Ala

          1733       1742       1751       1760       1769       1778
   ATG GAA AAA GAT CCC AAC AGC GTG GTC ATT GGC GCT GGC GCC TCT ATT GGC AGC
   Met Glu Lys Asp Pro Asn Ser Val Val Ile Gly Ala Gly Ala Ser Ile Gly Ser

          1787       1796       1805       1814       1823       1832
   CAG GAC TGG ATG AAG TCG GCA TTG CTG GCG CAA AAG GCG AAC GTC GAC CCG CAC
   Gln Asp Trp Met Lys Ser Ala Leu Leu Ala Gln Lys Ala Asn Val Asp Pro His

          1841       1850       1859       1868       1877       1886
   AAG ATG CGC TAC GTT GCC TTT GAG GGC GGC GGC GAG CCG GTG ACG GCA TTA ATG
   Lys Met Arg Tyr Val Ala Phe Glu Gly Gly Gly Glu Pro Val Thr Ala Leu Met

          1895       1904       1913       1922       1931       1940
   GGC AAC CAT GTT CAG GTT GTC TCC GGC GAT CTC AGT GAA ATG GTG CCT TAT CTG
   Gly Asn His Val Gln Val Val Ser Gly Asp Leu Ser Glu Met Val Pro Tyr Leu

          1949       1958       1967       1976       1985       1994
   GGC GGC GAC AAA ATC CGC GTG CTT GCC GTC TTT TCA GAA AAT CGT CTG CCG GGC
   Gly Gly Asp Lys Ile Arg Val Leu Ala Val Phe Ser Glu Asn Arg Leu Pro Gly
```

FIGURE 6B

```
      2003       2012       2021       2030       2039       2048
CAG CTT GCC AAT ATC CCT ACC GCT AAA GAA CAG GGG TAC GAC CTG GTG TGG CCG
Gln Leu Ala Asn Ile Pro Thr Ala Lys Glu Gln Gly Tyr Asp Leu Val Trp Pro

      2057       2066       2075       2084       2093       2102
ATT ATT CGC GGC TTC TAC GTC GGG CCC AAA GTC AGC GAT GCC GAT TAC CAG TGG
Ile Ile Arg Gly Phe Tyr Val Gly Pro Lys Val Ser Asp Ala Asp Tyr Gln Trp

      2111       2120       2129       2138       2147       2156
TGG GTG GAT ACC TTC AAG AAG CTC CAG CAA ACC GAC GAG TTT AAA AAG CAG CGC
Trp Val Asp Thr Phe Lys Lys Leu Gln Gln Thr Asp Glu Phe Lys Lys Gln Arg

      2165       2174       2183       2192       2201       2210
GAT CTG CGC GGC CTG TTT GAG TTC GAC ATG ACC GGC CAG CAG CTC GAT GAC TAC
Asp Leu Arg Gly Leu Phe Glu Phe Asp Met Thr Gly Gln Gln Leu Asp Asp Tyr

      2219       2228       2237       2246       2255       2264
GTG AAA AAA CAG GTT ACT GAT TAC CGT GAA CAG GCG AAA GCC TTC GGA CTC GCG
Val Lys Lys Gln Val Thr Asp Tyr Arg Glu Gln Ala Lys Ala Phe Gly Leu Ala


AAA TAA 3'
Lys ***
```

FIGURE 7A

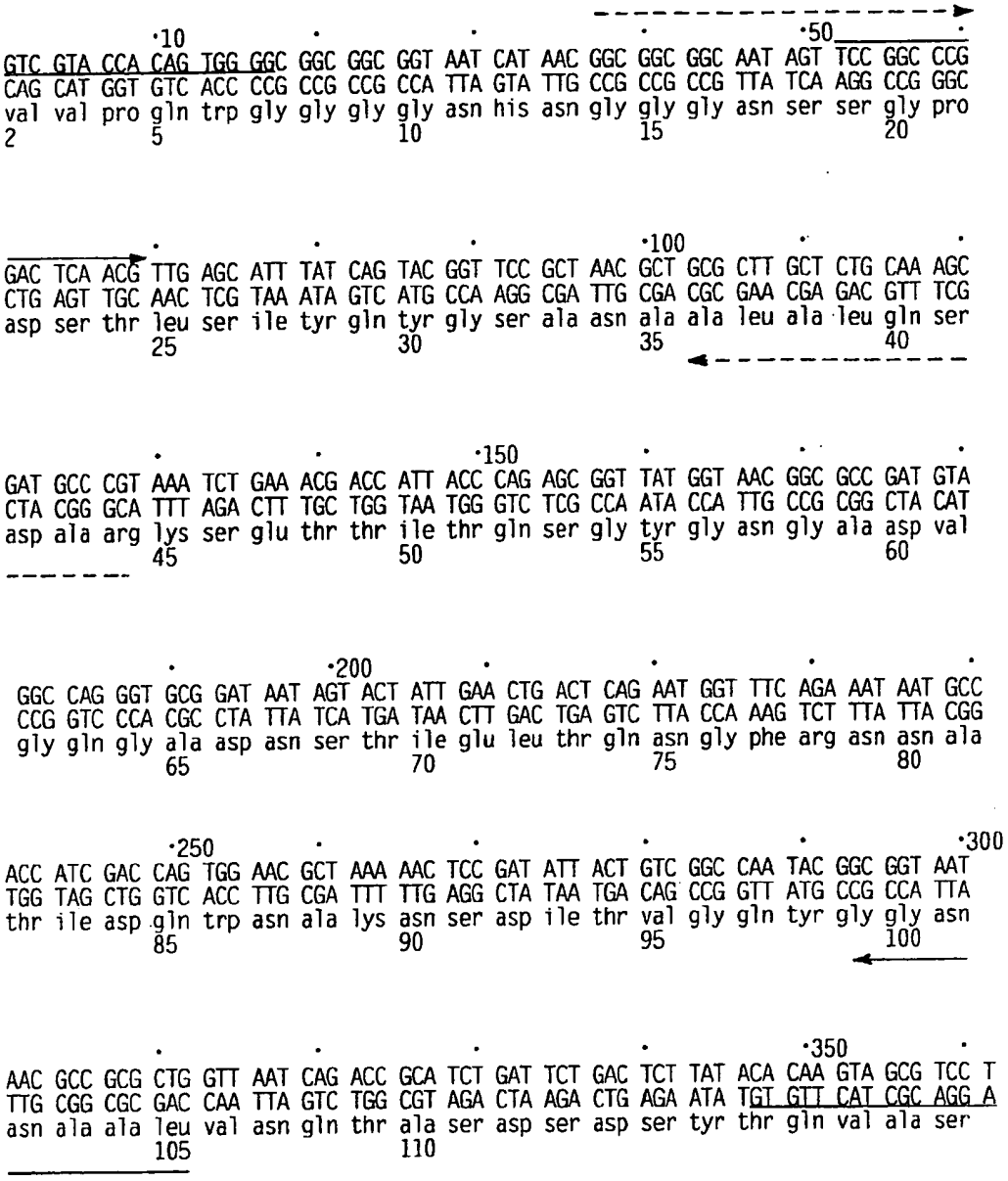
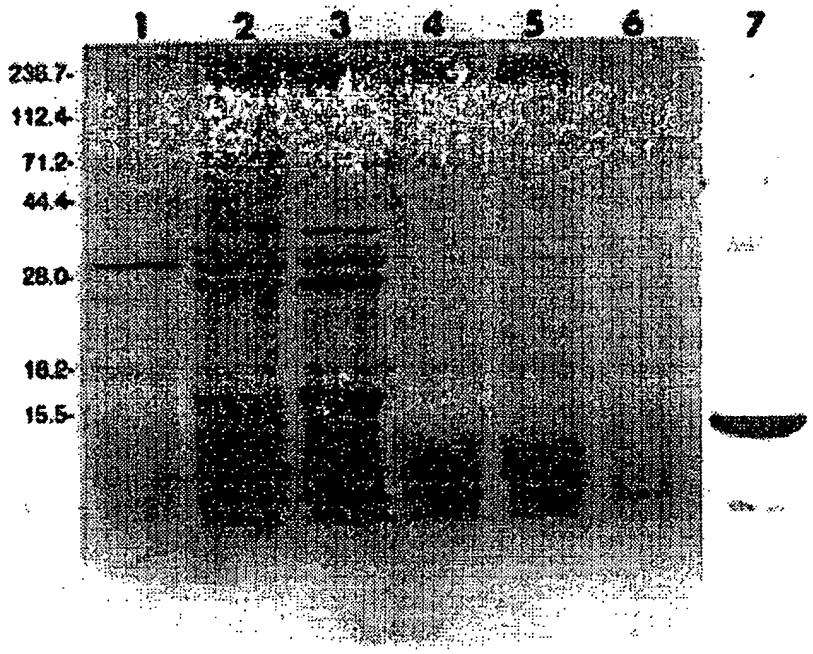

·10 ·50
GTC GTA CCA CAG TGG GGC GGC GGC GGT AAT CAT AAC GGC GGC GGC AAT AGT TCC GGC CCG
CAG CAT GGT GTC ACC CCG CCG CCG CCA TTA GTA TTG CCG CCG CCG TTA TCA AGG CCG GGC
val val pro gln trp gly gly gly gly asn his asn gly gly gly asn ser ser gly pro
2 5 10 15 20
·100
GAC TCA ACG TTG AGC ATT TAT CAG TAC GGT TCC GCT AAC GCT GCG CTT GCT CTG CAA AGC
CTG AGT TGC AAC TCG TAA ATA GTC ATG CCA AGG CGA TTG CGA CGC GAA CGA GAC GTT TCG
asp ser thr leu ser ile tyr gln tyr gly ser ala asn ala ala leu ala leu gln ser
25 30 35 40
·150
GAT GCC CGT AAA TCT GAA ACG ACC ATT ACC CAG AGC GGT TAT GGT AAC GGC GCC GAT GTA
CTA CGG GCA TTT AGA CTT TGC TGG TAA TGG GTC TCG CCA ATA CCA TTG CCG CGG CTA CAT
asp ala arg lys ser glu thr thr ile thr gln ser gly tyr gly asn gly ala asp val
45 50 55 60
·200
GGC CAG GGT GCG GAT AAT AGT ACT ATT GAA CTG ACT CAG AAT GGT TTC AGA AAT AAT GCC
CCG GTC CCA CGC CTA TTA TCA TGA TAA CTT GAC TGA GTC TTA CCA AAG TCT TTA TTA CGG
gly gln gly ala asp asn ser thr ile glu leu thr gln asn gly phe arg asn asn ala
65 70 75 80
·250 ·300
ACC ATC GAC CAG TGG AAC GCT AAA AAC TCC GAT ATT ACT GTC GGC CAA TAC GGC GGT AAT
TGG TAG CTG GTC ACC TTG CGA TTT TTG AGG CTA TAA TGA CAG CCG GTT ATG CCG CCA TTA
thr ile asp gln trp asn ala lys asn ser asp ile thr val gly gln tyr gly gly asn
85 90 95 100
·350
AAC GCC GCG CTG GTT AAT CAG ACC GCA TCT GAT TCT GAC TCT TAT ACA CAA GTA GCG TCC T
TTG CGG CGC GAC CAA TTA GTC TGG CGT AGA CTA AGA CTG AGA ATA TGT GTT CAT CGC AGG A
asn ala ala leu val asn gln thr ala ser asp ser asp ser tyr thr gln val ala ser
105 110

FIGURE 7B

```
1/1                                                    31/11
ATG AAA CTT TTA AAA GTG GCA GCA TTC GCA GCA ATC GTA GTT TCT GGC AGT GCT CTG GCT
Met lys leu leu lys val ala ala phe ala ala ile val val ser gly ser ala leu ala

61/21                                                  91/31
GGC GTC GTT CCA CAA TGG GGC GGC GGC GGT AAT CAT AAC GGC GGC GGC AAT AGT TCC GGC
gly val val pro gln trp gly gly gly gly asn his asn gly gly gly asn ser ser gly

121/41                                                 151/51
CCG GAC TCA ACG TTG AGC ATT TAT CAG TAC GGT TCC GCT AAC GCT GCG CTT GCT CTG CAA
pro asp ser thr leu ser ile tyr gln tyr gly ser ala asn ala ala leu ala leu gln

181/61                                                 211/71
AGC GAT GCC CGT AAA TCT GAA ACG ACC ATT ACC CAG AGC GGT TAT GGT AAC GGC GCC GAT
ser asp ala arg lys ser glu thr thr ile thr gln ser gly tyr gly asn gly ala asp

241/81                                                 271/91
GTA GGC CAG GGT GCG GAT AAT AGT ACT ATT GAA CTG ACT CAG AAT GGT TTC AGA AAT AAT
val gly gln gly ala asp asn ser thr ile glu leu thr gln asn gly phe arg asn asn

301/101                                                331/111
GCC ACC ATC GAC CAG TGG AAC GCT AAA AAC TCC GAT ATT ACT GTC GGC CAA TAC GGC GGT
ala thr ile asp gln trp asn ala lys asn ser asp ile thr val gly gln tyr gly gly

361/121                                                391/131
AAT AAC GCC GCG CTG GTT AAT CAG ACC GCA TCT GAT TCC AGC GTA ATG GTG CGT CAG GTT
asn asn ala ala leu val asn gln thr ala ser asp ser ser val met val arg gln val

421/141                                                451/151
GGT TTT GGC AAC AAC GCC ACG GCT AAC CAG TAT TAA
gly phe gly asn asn ala thr ala asn gln tyr OCH
```

FIG. 8A


1
2
3
4
5
6
7
238.7
112.4
71.2
44.4
28.0
18.2
15.5

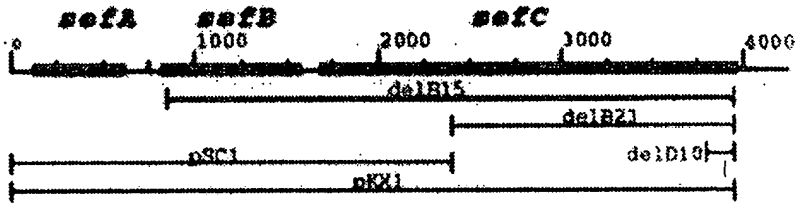

sofA
sofB
sofC
0
1000
2000
3000
4000
delB15
delB21
pSC1
delD10
pKX1

FIG. 8B

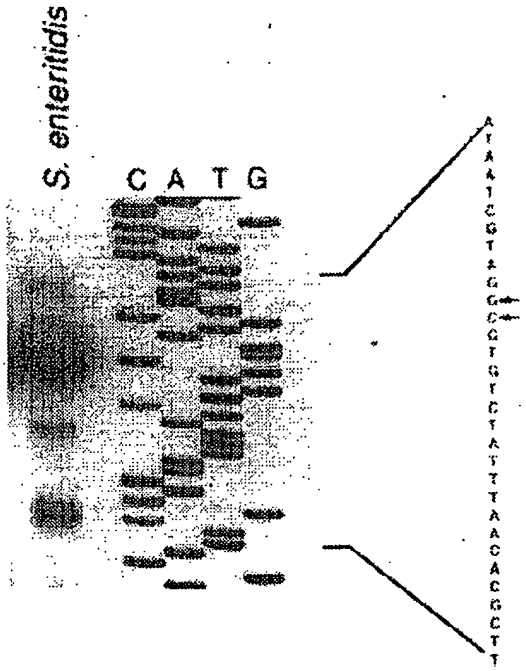


FIG. 9

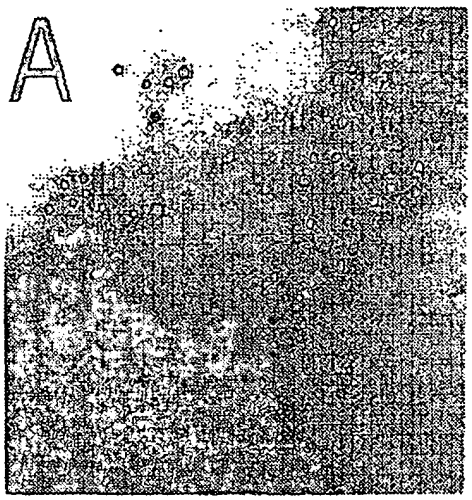


FIG. 10A

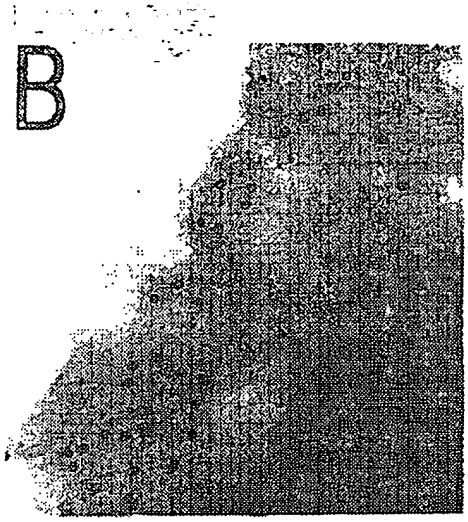


FIG. 10B

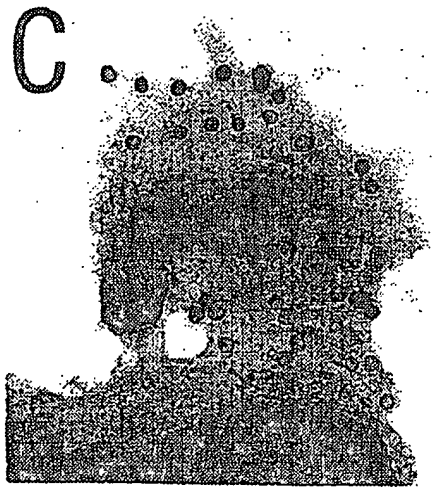


FIG. 10C